LWRfirstpage

Probability with Measure

Dr Nic Freeman

May 5, 2022

Contents

Chapter 0

Introduction

0.1 Organization

0.1.1 Syllabus

These notes are for three courses: MAS350, MAS451 and the Spring semester of MAS6352.

Some sections of the course are included in MAS451/6352 but not in MAS350. These sections are marked with a (Δ) symbol. We will not cover these sections in lectures. Students taking MAS451/6352 should study these sections independently.

Some parts of the notes are marked with a ($\star$) symbol, which means they are off-syllabus. These are often cases where detailed connections can be made to and from other parts of mathematics.

0.1.2 Problem sheets

The exercises are divided up according to the chapters of the course. Some exercises are marked as 'challenge questions' – these are intended to offer a serious, time consuming challenge to the best students.

Aside from challenge questions, it is expected that students will attempt all exercises (for the version of the course they are taking) and review their own solutions using the typed solutions provided at the end of these notes.

At three points during each semester, an assignment of additional exercises will be set. About one week later, a mark scheme will be posted, and you should self-mark your solutions.

0.1.3 Examination

The course will be examined in the summer sitting. Parts of the course marked with a (Δ) are examinable for MAS451/6352 but not for MAS350. Parts of the course marked with a ($\star$) will not be examined (for everyone).

0.1.4 Website

Further information, including the timetable, can be found on

`http://nicfreeman.staff.shef.ac.uk/MASx50/`.

0.2 Preliminaries

This section contains lots of definitions, from earlier courses, that we will use in MAS350. Most of the material here should be familiar to you. There may be one or two minor extensions of ideas you have seen before.

1. *Set Theory.*

   Let S be a set and $A, B, C, \ldots$ be subsets.

   A^c is the complement of A in S so that

   $$A^c = \{x \in S; x \notin A\}.$$

   Union $A \cup B = \{x \in S; x \in A \text{ or } x \in B\}$.

   Intersection $A \cap B = \{x \in S; x \in A \text{ and } x \in B\}$.

   Set theoretic difference: $A - B = A \cap B^c$.

   We have finite and infinite unions and intersections so if $A_1, A_2, \ldots, A_n$ are subsets of S.

   $$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n.$$
   $$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

   We will also need *infinite* unions and intersections. So let (A_n) be a sequence of subsets in S.

   Let $x \in S$. We say that $x \in \bigcup_{i=1}^{\infty} A_i$ if $x \in A_i$ for at least one value of i. We say that $x \in \bigcap_{i=1}^{\infty} A_i$ if $x \in A_i$ for all values of i.

   Note that *de Morgan's laws* of set algebra hold in this context:

   $$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c.$$
   $$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c.$$

   The *Cartesian product* $S \times T$ of sets S and T is

   $$S \times T = \{(s,t); s \in s, t \in T\}.$$

2. *Sets of Numbers*

   - Natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.
   - Non-negative integers $\mathbb{Z}_+ = \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$.
   - Integers $\mathbb{Z}$.
   - Rational numbers $\mathbb{Q}$.
   - Real numbers $\mathbb{R}$.

- Complex numbers $\mathbb{C}$.

A set X is *countable* if there exists an injection between X and $\mathbb{N}$. A set is *uncountable* if it fails to be countable. $\mathbb{N}, \mathbb{Z}_+, \mathbb{Z}$ and $\mathbb{Q}$ are countable. $\mathbb{R}$ and $\mathbb{C}$ are uncountable. All finite sets are countable.

3. *Images and Preimages.*

Suppose that S_1 and S_2 are two sets and that $f : S_1 \to S_2$ is a mapping (or function). Suppose that $A \subseteq S_1$. The *image* of A under f is the set $f(A) \subseteq S_2$ defined by

$$f(A) = \{y \in S_2; y = f(x) \text{ for some } x \in S_1\}.$$

If $B \subseteq S_2$ the *inverse image* of B under f is the set $f^{-1}(B) \subseteq S_1$ defined by

$$f^{-1}(B) = \{x \in S_1; f(x) \in B\}.$$

Note that $f^{-1}(B)$ makes sense irrespective of whether the mapping f is invertible.

Key properties are, with $A, A_1, A_2 \subseteq S_1$ and $B, B_1, B_2 \subseteq S_2$:

$$\begin{aligned}
f^{-1}(B_1 \cup B_2) &= f^{-1}(B_1) \cup f^{-1}(B_2), \\
f^{-1}(B_1 \cap B_2) &= f^{-1}(B_1) \cap f^{-1}(B_2), \\
f^{-1}(A^c) &= f^{-1}(A)^c, \\
f(A_1 \cup A_2) &= f(A_1) \cup f(A_2), \\
f(A_1 \cap A_2) &\subseteq f(A_1) \cap f(A_2), \\
f(f^{-1}(B)) &\subseteq B \\
(f \circ g)^{-1}(A) &= g^{-1}(f^{-1}(A)) \\
A &\subseteq f^{-1}(f(A)), \\
A \subseteq B \quad \Rightarrow \quad f^{-1}(A) &\subseteq f^{-1}(B).
\end{aligned}$$

4. *Extended Real Numbers*

We will often find it convenient to work with ∞ and $-\infty$. These are *not* real numbers, but we find it convenient to treat them a bit like real numbers. To do so we specify the extra arithmetic rules, for all $x \in \mathbb{R}$,

$$\infty + x = x + \infty = \infty,$$

$$x - \infty = -\infty + x = -\infty,$$

$$\infty.x = x.\infty = \infty \text{ for } x > 0,$$

$$\infty.x = x.\infty = -\infty \text{ for } x < 0,$$

$$\infty.0 = 0.\infty = 0.$$

Note that $\infty - \infty$, $\infty.\infty$ and $\frac{\infty}{\infty}$ are undefined. We also specify that, for all $x \in \mathbb{R}$,

$$-\infty < x < \infty.$$

We write $\mathbb{R}^* = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$, which is known as the *extended* real numbers.

5. *Analysis.*

- sup and inf. If A is a bounded set of real numbers, we write $\sup(A)$ and $\inf(A)$ for the real numbers that are their least upper bounds and greatest lower bounds (respectively.) If A fails to be bounded above, we write $\sup(A) = \infty$ and if A fails to be bounded below we write $\inf(A) = -\infty$. Note that $\inf(A) = -\sup(-A)$ where $-A = \{-x; x \in A\}$. If $f : S \to \mathbb{R}$ is a mapping, we write $\sup_{x\in S} f(x) = \sup\{f(x); x \in S\}$. A very useful inequality is
$$\sup_{x\in S} |f(x) + g(x)| \leq \sup_{x\in S} |f(x)| + \sup_{x\in S} |g(x)|.$$
- Sequences and Limits. Let $(a_n) = (a_1, a_2, a_3, \ldots)$ be a sequence of real numbers. It *converges* to the real number a if given any $\epsilon > 0$ there exists a natural number N so that whenever $n > N$ we have $|a - a_n| < \epsilon$. We then write $a = \lim_{n\to\infty} a_n$.

  A sequence (a_n) which is *monotonic increasing* (i.e. $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$) and *bounded above* (i.e. there exists $K > 0$ so that $a_n \leq K$ for all $n \in \mathbb{N}$) converges to $\sup_{n\in\mathbb{N}} a_n$.

  A sequence (a_n) which is *monotonic decreasing* (i.e. $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$) and *bounded below* (i.e. there exists $L > 0$ so that $a_n \geq L$ for all $n \in \mathbb{N}$) converges to $\inf_{n\in\mathbb{N}} a_n$.

  A *subsequence* of a sequence (a_n) is itself a sequence of the form (a_{n_k}) where $n_{k_1} < n_{k_2}$ when $k_1 < k_2$.
- Series. If the sequence (s_n) converges to a limit s where $s_n = a_1 + a_2 + \cdots + a_n$ we write $s = \sum_{n=1}^{\infty} a_n$ and call it the *sum of the series.* If each $a_n \geq 0$ then the sequence (s_n) is either convergent to a limit or properly divergent to infinity. In the latter case we write $s = \infty$ and interpret this in the sense of extended real numbers.
- Continuity. A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at $a \in \mathbb{R}$ if given any $\epsilon > 0$ there exists $\delta > 0$ so that $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon$. Equivalently f is continuous at a if given any sequence (a_n) that converges to a, the sequence $(f(a_n))$ converges to $f(a)$.

  f is a *continuous function* if it is continuous at every $a \in \mathbb{R}$.

Chapter 1

Measure Spaces

easure_spaces

1.1 What is measure theory?

sec:intro

Measure theory is the abstract mathematical theory that underlies all models of measurement of 'size' in the real world. This includes measurement of length, area and volume, weight and mass, and also of chance and probability. Measure theory is a branch of pure mathematics, in particular of analysis, but it plays key roles in both calculus and statistical modelling. This is because measure theory provides the foundation of both the modern theory of integration and of the modern theory of probability.

Suppose that we wish to measure the lengths of several line segments. We represent these as closed intervals of the real number line $\mathbb{R}$ so a typical line segment is $[a, b]$ where $b > a$. We all agree that its length is $b - a$. We write this as

$$m([a, b]) = b - a$$

and interpret this as telling us that the measure m of length of the line segment $[a, b]$ is the number $b - a$. We might also agree that if $[a_1, b_1]$ and $[a_2, b_2]$ are two non-overlapping line segments and we want to measure their combined length then we want to apply m to the set-theoretic union $[a_1, b_1] \cup [a_2, b_2]$ and

$$m([a_1, b_1] \cup [a_2, b_2]) \quad = (b_2 - a_2) + (b_1 - a_1) = m([a_1, b_1]) + m([a_2, b_2]). \tag{1.1}$$

An isolated point c has zero length and so

$$m(\{c\}) = 0.$$

If we consider the whole real line in its entirety then it has infinite length, i.e.

$$m(\mathbb{R}) = \infty.$$

The key point here is that, if we try to abstract the notion of a 'measure of length, then we should regard it as a mapping m defined on subsets of the real line, that takes values in the extended non-negative real numbers $[0, \infty]$.

We might wonder why there is any mathematical difficulty involved here, since it appears that we can easily agree on how how long a line is. The problem is that subsets of $\mathbb{R}$ may arise naturally and still be rather complicated.

ex:cantor **Example 1.1.1 (The Cantor Set)** Let $C_0 = [0,1]$. Given C_n, define C_{n+1} by taking each sub-interval of C_n, cutting this sub-interval into three parts of equal length and removing the open interval corresponding to the middle third. So, for each n, C_n is a set of 2^n closed intervals each of length $(\frac{1}{3})^n$.

Let $C = \bigcap_{n=0}^{\infty} C_n$. Clearly $C_{n+1} \subseteq C_n$, so this is a decreasing sequence of sets, and C is precisely the points that 'never end up in the middle thirds'. For example, $0 \in C_n$ and $\frac{1}{3} \in C_n$.

The total length of the intervals in C_n is $2^n(\frac{1}{3})^n = (\frac{2}{3})^n$, which tends to zero as $n \to \infty$. This suggests C should have 'length' zero, but how can we make this intuition into rigorous mathematics?

The Cantor set is a 'fractal', which is a general term for any shape with very detailed structure. It is somewhat contrived – in fact, it was first introduced precisely as a contrived example of an odd looking shape that appeared to exist within the real line, but with no obvious purpose. Today, we know that fractal-like objects appear frequently within nature, which means that we also need to deal with them within our theory of measure.

1.2 Sigma Fields

We need to be more ambitious that just measuring the length of intervals of $\mathbb{R}$. More generally, we want to work with a function m such that the map $A \mapsto m(A)$ corresponds to our intuitive idea of measuring how 'big' the object A is. Length is one example of this, 'weight' and 'volume' are other examples. The function m will be known as *a measure*, and we say that m *measures* the length/volume/size/weight/etc of A.

To do this rigorously, the first question we must answer is: which objects are we going to measure? This question has a reasonably straightforward answer. We are going to take a set S, and we are going to 'measure' subsets A of the set S. Note that at this stage we don't specify *what property* of A we are going to measure. We might measure length, or volume, or some other property that might be more difficult to express in words.

However, there is a caveat. In many cases, particularly if the set S is very large (such as $\mathbb{R}$ itself, which is uncountable) we will not be able to measure the size of *every* subset of S. The reasons for this caveat are difficult, and we will come to them in Section 1.6. Instead, we do the next best thing. We specify precisely *which* subsets of S we are going to measure.

f:sigma_field **Definition 1.2.1** Let S be a set. A *σ-field* on S is a set Σ, such that each $A \in \Sigma$ is a subset of S, satisfying the following properties:

(S1) $\emptyset \in \Sigma$ and $S \in \Sigma$.

(S2) If $A \in \Sigma$ then $A^c \in \Sigma$.

(S3) If $(A_n)_{n\in\mathbb{N}}$ is a sequence of sets with $A_n \in \Sigma$ for all $n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Definition 1.2.2 Given a σ-field Σ, a set $A \in \Sigma$ is said to be *measurable* with respect to Σ. We will often shorten this to 'Σ-measurable', or simply 'measurable' if the context makes clear which Σ is meant.

The purpose of (S1)-(S3) is to capture some of our intuition on what it means 'to measure'. Let us go through them carefully. The first part of (S1) says that we should be able to measure a set $\emptyset$ that is empty (and, when the time comes, we will force $\emptyset$ to have measure zero). Property (S2) is a statement that if we are going to be able to measure A, we also want to be able to measure its complement $A^c = S \setminus A$. This is very natural from a physical point of view: if you have a 1kg bag of flour and you take 450g out, then you expect to be able to measure how much flour you have left. The complement of $\emptyset$ is $\emptyset^c = S \setminus \emptyset = S$, so this means we also need to be able to measure S itself, thus leading us to the second half of (S1).

Property (S3) is a bit more subtle. Firstly note that if we can measure A and B then it is reasonable (again, think flour) to want to measure their union $A \cup B$. Similarly, if we can measure $A_1, \dots, A_n$ then it is reasonable to want to measure their union $\cup_{i=1}^{n} A_n$. However, we can't stop here. We need our theory of measure to handle infinite objects, like the interval $[0, 2]$ which contains infinitely many elements (even though only has length 2!). For this reason we also allow countable unions, of the form $\cup_{n=1}^{\infty} \dots$ in (S3).

Remark 1.2.3 *We cannot 'upgrade' to allowing uncountable unions in (S3). Doing so would, unfortunately, break our entire theory of measure, for a reason that we cannot easily see, yet. We will discuss this point further in Section 1.6.*

Remark 1.2.4 *The term 'σ-algebra' is used by some books, with the same meaning as 'σ-field'. I prefer σ-field, you may use either.*

Let us briefly note a few properties of σ-fields, which build on the properties (S1)-(S3).

- We have seen in (S3) that Σ is closed under countably infinite unions, meaning that taking a countable unions of sets in Σ gives back a set in Σ. The same is true of finite unions. To see this let $A_1, \ldots, A_n \in \Sigma$ and define $A_i = \emptyset$ for $i > n$. By (S1) we have $A_i \in \Sigma$ for all $i \in \mathbb{N}$. Note that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$, and thus by (S3) we have $\bigcup_{i=1}^{n} A_i \in \Sigma$.

- Σ is also closed under countably infinite intersections. To see this we can use the laws of set algebra to write $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$, and then apply (S2) and (S3) to the right hand side. By the same ideas as above, Σ is also closed under finite intersections.

- Σ is also closed under set theoretic differences. To see this note that $A \setminus B = A \cap B^c$, and apply (S2) along with closure under intersections to the the right hand side.

We can summarise the above properties as follows: if we have a σ-field Σ, and sets $A_1, A_2, \ldots \in \Sigma$, then applying any finite or countable number of set operations to the A_i will simply give us back another set in Σ. We call this fact 'closure under countable set operations'. We will use it repeatedly throughout the course.

Definition 1.2.5 A pair (S, Σ) where S is a set and Σ is a σ-field of subsets of S is called a *measurable space*.

Given a set S, there are typically are typically many possible choices of Σ. The choice of Σ is determined by what it is that we want to measure.

1.2.1 Examples of σ-fields

The following examples are all σ-fields.

1. For any set S, the power set $\mathcal{P}(S)$ is a σ-field. Recall that $\mathcal{P}(S)$ is the set of all subsets of S, so (S1)-(S3) are automatically satisfied.

2. For any set S, $\Sigma = \{\emptyset, S\}$ is a σ-field, called the *trivial* σ-field.

3. If S is any set and $A \subset S$ then $\Sigma = \{\emptyset, A, A^c, S\}$ is a σ-field. Checking (S1)-(S3) in this case is left for you.

4. Similarly, if $A, B \subset S$ then $\Sigma = \{\emptyset, A, B, A \cup B, (A \cup B)^c, A \cap B, (A \cap B)^c, A \setminus B, (A \setminus B)^c, B \setminus A, (B \setminus A)^c, (A \cup B) \setminus (A \cap B), ((A \cup B) \setminus (A \cap B))^c, A \cup B^c, A^c \cup B, S\}$ is a σ-field. I suggest not checking this one.

I hope this is a convincing demonstration that we cannot hope to simply write down σ-fields, for the most part. Instead we need a tool for constructing them, without needing to write them down. This is done as follows.

gma_intersect **Lemma 1.2.6** *Let I be any set and for each $i \in I$ let Σ_i be a σ-field on S. Then*

$$\Sigma = \bigcap_{i \in I} \Sigma_i \tag{1.2}$$

{eq:sigma_in

is a σ-field on S.

PROOF: We check the three conditions of Definition 1.2.1 for $\mathcal{F}$.

(S1) Since each Σ_i is a σ-field, we have $\emptyset \in \Sigma_i$. Hence $\emptyset \in \cap_i \Sigma_i$. Similarly, $S \in \Sigma$.

(S2) If $A \in \Sigma = \cap_i \mathcal{F}_i$ then $A \in \Sigma_i$ for each i. Since each Σ_i is a σ-field, $S \setminus A \in \Sigma_i$ for each i. Hence $S \setminus A \in \cap_i \Sigma_i$.

(S3) If $A_j \in \Sigma$ for all j, then $A_j \in \Sigma_i$ for all i and j. Since each Σ_i is a σ-field, $\cup_j A_j \in \Sigma_i$ for all i. Hence $\cup_j A_j \in \cap_i \Sigma_i$. ■

Example 1.2.7 Thanks to Lemma 1.2.6 we can construct a σ-field by making a statement along the lines of

> *"Let Σ be the smallest σ-field on $\mathbb{R}$ containing all the open intervals."*

By this statement we mean: let Σ be intersection of all the σ-fields on $\mathbb{R}$ that contain all of the open intervals, in the style of equation (1.2). We know that at least one σ-field exists with this property, namely $\mathcal{P}(\mathbb{R})$. Therefore Lemma 1.2.6 applies, and tells that Σ is indeed a σ-field. The σ-field resulting from this example is very special. It is known as the Borel σ-field on $\mathbb{R}$, and it is much smaller than $\mathcal{P}(\mathbb{R})$. We will introduce it formally in Definition 1.4.2, and study it in Section 1.4.

1.3 Measure

The next question we need to ask is: what does it mean to measure an object? We want a general framework that we can use for concepts such as length, weight and volume. From the last section, we know that we are looking for a function $m : \Sigma \to [0,\infty]$ where Σ will be an appropriately chosen σ-field.

def:measure **Definition 1.3.1** Let (S,Σ) be a measurable space. A mapping $m : \Sigma \to [0,\infty]$ is known as a *measure* if it satisfies

(M1) $m(\emptyset) = 0$.

(M2) If $(A_n)_{n\in\mathbb{N}}$ is a sequence of sets where each $A_n \in \Sigma$ and if these sets are pairwise disjoint (meaning that $A_n \cap A_m = \emptyset$ if $m \neq n$) then

$$m\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n).$$

Note that (M2) relies on property (S3), to make sure that $\bigcup_{n=1}^{\infty} A_n \in \Sigma$. Property (M2) is often known as σ-additivity. Crucially, it will allow us to take limits in ways that involve measures, thanks to the fact that (M2) considers a (countably) infinite sequence of sets (A_n). Limits are how we rigorously justify that approximations work – consequently we need them, if we are to create a theory that will, ultimately, be useful to experimentalists and modellers.

Property (M2) encapsulates the idea that if we take a collection of objects, then their total measure should equal to the sum of their individual measures – providing they don't overlap with each other. For example, we might take 1kg of flour and divide it into 3 piles weighing 100g, 250g and 650g. We could also imagine dividing our 1kg of flour into an infinite sequence of piles, with sizes 500g, 250g, 125g, 67.5g, ..., that sum (as an infinite series) to 1kg.

Property (M1) is much less remarkable. It simply states that the empty set has zero measure. This represents our feeling that an empty region of space has zero length/weight/volume/etc.

Definition 1.3.2 A triplet (S,Σ,m) where S is a set, Σ is a σ-field on S, and $m : \Sigma \to [0,\infty]$ is a measure is known as a *measure space*.

Definition 1.3.3 The extended real number $m(S)$ is called the *total mass* of m. The measure m is said to be *finite* if $m(S) < \infty$.

Let us now assume that (S,Σ,m) is a measure space, and record some useful properties of measures.

- If $A_1,\ldots,A_n \in \Sigma$ and are pairwise disjoint then

  $$m(A_1 \cup \ldots \cup A_n) = m(A_1) + \ldots + m(A_n).$$

  This is known as *finite additivity* of measures. We'll often think of it as part of (M2).

  To prove it we use the same idea on (M2) as we used, for σ-fields, on (S3). Define $A'_i = A_i$ for $i \leq n$ and $A'_i = \emptyset$ for $i > n$. By (M2) we have $m\left(\bigcup_{i=1}^{\infty} A'_i\right) = \sum_{i=1}^{\infty} m(A'_i)$. By (M1) we have $m(\emptyset) = 0$, so this reduces to $m(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} m(A_i)$.

- If $A, B \in \Sigma$ with $A \subseteq B$ then $m(A) \leq m(B)$. This property is known as the *monotonicity* property of measures.

  To prove it write B as the disjoint union $B = (B \setminus A) \cup A$ and then use that from part 1 we have $m(B) = m((B \setminus A) \cup A) = m(B \setminus A) + m(A)$. If $m(A)$ is finite we can subtract it from both sides, and obtain that $m(B \setminus A) = m(B) - m(A)$. However, this only works if $m(A)$ is finite!

- If $A, B \in \Sigma$ are arbitrary (i.e. not necessarily disjoint) then

$$m(A \cup B) + m(A \cap B) = m(A) + m(B). \tag{1.3}$$

{munion}

  The proof of this is Problem **1.3** part (a). Note that if $m(A \cap B) < \infty$ we have $m(A \cup B) = m(A) + m(B) - m(A \cap B)$, which you might recognize as similar to something you've seen before in probability.

1.3.1 Examples of measures

Here are three important first examples of measure spaces. We can't yet introduce examples based on length or volume; this will come later in the course.

1. **Counting Measure** Let S any set and take $\Sigma = \mathcal{P}(S)$. For each $A \subseteq S$ the counting measure $m = \#$ is given by

   $$\#(A) = \text{the number of elements in } A.$$

   I hope its intuitively obvious to you that this is a measure. We'll omit checking the details.

2. **Dirac Measure** This measure is named after the physicist Paul Dirac. Let (S, Σ) be an arbitrary measurable space and fix $x \in S$. The Dirac measure $m = \delta_x$ is defined by

   $$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

   Checking properties (M1) and (M2) in this case is left for you.

   A useful fact: if S is countable then we can write the counting measure $\#$ in terms of Dirac measures, as $\#(A) = \sum_{x \in S} \delta_x(A)$.

3. **Probability**

   Consider a finite set $S = \{x_1, \ldots, x_n\}$, which we'll call the *sample space* and call each of the x_i an *outcome*. Let Σ be the set of all subsets of S. Let $(p_i)_{i=1}^n$ be set of numbers in $[0, 1]$ such that $\sum_{i=1}^n p_i = 1$. For $A \in \Sigma$ we define a measure $m = \mathbb{P}$ by setting

   $$\mathbb{P}[A] = \sum_{i=1}^n p_i \delta_{x_i}(A). \tag{1.4}$$ {eq:prob_dir

   In words, to each outcome x_i we assign probability p_i, that is $\mathbb{P}[\{x_i\}] = p_i$. If a set A contains several outcomes, then its outcome is precisely the sum of their individual probabilities. Finding the probability of an event is just another kind of measuring!

   We could treat a countable set S similarly, with a countable sequence of p_i and a countable summation (i.e. an infinite series) in (1.4). Probability, however, mostly requires uncountable sample spaces (e.g. the normal distribution on the real line). In this case (1.4) breaks down completely, because there is no such thing as an uncountable sum. One of the outcomes of this course will be a rigorous basis for probability theory with uncountable sample spaces.

   In general, a measure m is said to be a *probability* measure if its total mass is 1 i.e. $m(S) = 1$.

4. **Integration**

   In MAS221 you viewed Riemann integration as a way of calculating area – that is, measuring the area of two-dimensional shapes. You've probably also viewed various types of integrals as ways of calculating volumes, at some point. So, we should expect integration to fit naturally into our theory of measures.

   In Chapter 3 we will introduce *Lebesgue integration*. Lebesgue integration is 'the' modern theory of integration on which mathematical modelling now relies. We will see that Lebesgue integration interacts nicely with measure theory, whilst Riemann integration doesn't. In fact, Lebesgue integration will also be the key tool for setting up a rigorous basis for probability theory.

1.4 The Borel σ-field

In this section introduce another example of a measure space, which will represent the notion of measuring the 'length' of subsets of $\mathbb{R}$. For an interval $[a, b]$ it is clear that the length should be $b - a$, but as we saw in Section 1.1 for more complicated subsets of $\mathbb{R}$ the situation is not so clear.

Example 1.4.1 Consider, for example, the irrational numbers $\mathbb{I}$ and the rational numbers $\mathbb{Q}$. Both $\mathbb{I}$ and $\mathbb{Q}$ are found throughout $\mathbb{R}$, but they are both full of tiny holes. Can we find a meaningful way to decide what is the 'length' of $\mathbb{I}$ and $\mathbb{Q}$? In fact, we will see that we can – to come in Section 1.5. But in Section 1.6 we will also show that it is possible to construct subsets of $\mathbb{R}$ for which there is *no* meaningful idea of length.

In this section we take $S = \mathbb{R}$. The first question is: which σ-field should we use? The power set $\mathcal{P}(\mathbb{R})$ is too big, for reasons that we will make clear in Section 1.6. However, for practical purposes we do need our σ-field to contain all open and closed intervals, and also unions, intersections and complements of these. This provides a starting point.

Definition 1.4.2 The *Borel σ-field* of $\mathbb{R}$, denoted by $\mathcal{B}(\mathbb{R})$, is the smallest σ-field on $\mathbb{R}$ that contains all open intervals (a, b) where $-\infty \leq a < b \leq \infty$. Sets in $\mathcal{B}(\mathbb{R})$ are called *Borel sets*.

Note that $\mathcal{B}(\mathbb{R})$ also contains isolated points $\{a\}$ where $a \in \mathbb{R}$. To see this first observe that $(a, \infty) \in \mathcal{B}(\mathbb{R})$ and also $(-\infty, a) \in \mathcal{B}(\mathbb{R})$. Now by (S2), $(-\infty, a] = (a, \infty)^c \in \mathcal{B}(\mathbb{R})$ and $[a, \infty) = (-\infty, a)^c \in \mathcal{B}(\mathbb{R})$. Finally as σ-fields are closed under intersections, $\{a\} = [a, \infty) \cap (-\infty, a] \in \mathcal{B}(\mathbb{R})$. You can show that $\mathcal{B}(\mathbb{R})$ also contains all closed intervals – see Problem **1.7**. With open and closed intervals in hand, the closure of σ-fields under countable set operations gives us a way to construct a huge variety of Borel sets.

As a general rule, all 'sensible' subsets of $\mathbb{R}$ are Borel sets. We might hope to find some sort of formula for a general element of $\mathcal{B}(\mathbb{R})$, but this is not possible. Unless you deliberately set out to find a non-Borel subset of $\mathbb{R}$ you will never come across one – and even when you look for them it is hard work to find them, as we will see in Section 1.6.

Remark 1.4.3 *($\star$) The Borel σ-field $\mathcal{B}(S)$ can be defined on any set S for which there are subsets that can be called 'open' in a sensible way. In particular this works for all metric spaces. The most general type of S for which you can form $\mathcal{B}(S)$ is a topological space.*

1.5 Lebesgue Measure

sec:leb_meas

The measure that precisely captures the notion of length is called *Lebesgue measure* in honour of the French mathematician Henri Lebesgue (1875-1941), who founded the modern theory of integration. We will denote it by λ. First we need a definition.

Let $A \in \mathcal{B}(\mathbb{R})$ be arbitrary. A *covering* of A is a finite or countable collection of open intervals $\{(a_n, b_n), n \in \mathbb{N}\}$ so that

$$A \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n). \tag{1.5}$$

{eq:covering}

def:leb_meas

Definition 1.5.1 Let $\mathcal{C}_A$ be the set of all coverings of the set $A \in \mathcal{B}(\mathbb{R})$. The *Lebesgue measure* λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined by the formula:

$$\lambda(A) = \inf_{\mathcal{C}_A} \sum_{n=1}^{\infty} (b_n - a_n), \tag{1.6}$$

{eq:leb}

where the inf is taken over all possible coverings of A, with notation as in (1.5)

It would take a long time to prove that λ really is a measure, and it wouldn't help us understand λ any better if we did it, so we'll omit that from the course. For the proof, see the standard text books e.g. Cohn, Schilling or Tao.

Let's check that the Definition 1.5.1 agrees with some of our intuitive ideas about length.

(L1) If $A = (a, b)$ then $\lambda((a, b)) = b - a$ as expected, since (a, b) is a covering of itself and any other cover will have greater length.

(L2) If $A = \{a\}$ then $\lambda(\{a\}) = 0$. To see this, choose any $\epsilon > 0$. Then $(a - \epsilon/2, a + \epsilon/2)$ is a cover of a and so $\lambda(\{a\}) \leq (a + \epsilon/2) - (a - \epsilon/2) = \epsilon$. But ϵ is arbitrary and so we conclude that $\lambda(\{a\}) = 0$.

(L3) Combining (L2) with (M2), we deduce that for $a < b$,

$$\lambda([a, b)) = \lambda(\{a\} \cup (a, b)) = \lambda(\{a\}) + \lambda((a, b)) = b - a.$$

Similarly, $\lambda([a, b]) = \lambda((a, b]) = b - a$.

(L4) If $A = [0, \infty)$, write $A = \bigcup_{n=1}^{\infty} [n - 1, n)$. Then by (M2) we obtain $\lambda([0, \infty)) = \sum_{n=1}^{\infty} 1 = \infty$. By a similar argument, $\lambda((-\infty, 0)) = \infty$ and so $\lambda(\mathbb{R}) = \lambda((-\infty, 0)) + \lambda([0, \infty)) = \infty$.

(L5) If $A \in \mathcal{B}(\mathbb{R})$, and for some $x \in \mathbb{R}$ we define $A_x = \{x + a\,;\, a \in A\}$, then $\lambda(A) = \lambda(A_x)$.

In words, if we take a set A and translate it (by x), we do not change its measure. We'll often refer to this property as the *translation invariance* of Lebesgue measure. It is easily seen from (1.6), because any cover of A can be translated by x to be a cover of A_x.

Example 1.5.2 In simple practical examples on Lebesgue measure, it is almost always best not to try to use (1.6) directly, but to just apply the properties listed above.

For example, to find $\lambda\left((-3, 10) \setminus (-1, 4)\right)$, use (L3) and (M2) to obtain

$$\begin{aligned}
\lambda\left((-3, 10) \setminus (-1, 4)\right) &= \lambda\left((-3, -1] \cup [4, 10)\right) \\
&= \lambda((-3, -1]) + \lambda([4, 10)) \\
&= ((-1) - (-3)) + (10 - 4) = 8
\end{aligned}$$

_restriction **Remark 1.5.3** *If I is a closed interval (or in fact any Borel set) in $\mathbb{R}$ we can similarly define $\mathcal{B}(I)$, the Borel σ-field of I, to be the smallest σ-field containing all open intervals in I. In fact, it holds that $\mathcal{B}(I) = \{B \cap I\,;\, B \in \mathcal{B}(\mathbb{R})\}$. The Lebesgue measure λ_I on $(I, \mathcal{B}(I))$ is obtained by restricting the sets A in (1.6) to be in $\mathcal{B}(I)$. It can be seen that for $A \subseteq I$ we have $\lambda_I(A) = \lambda(A)$. We won't include a proof of these claims.*

Sets of measure zero play an important role in measure theory, and in probability. We'll explore this in future sections. For now, here are some interesting examples of quite "large" sets that have Lebesgue measure zero

Lemma 1.5.4 *Let $A \subset \mathbb{R}$ be countable. Then $\lambda(A) = 0$.*

PROOF: Write $A = \{a_1, a_2, \ldots\} = \bigcup_{n=1}^{\infty}\{a_n\}$. Since A is an infinite union of point sets, it is in $\mathcal{B}(\mathbb{R})$. Then, using (M2) and (L2)

$$\lambda(A) = \lambda\left(\bigcup_{n=1}^{\infty}\{a_n\}\right) = \sum_{n=1}^{\infty}\lambda(\{a_n\}) = 0.$$

■

It follows that

$$\lambda(\mathbb{N}) = \lambda(\mathbb{Z}) = \lambda(\mathbb{Q}) = 0.$$

Further, for any $A \in \mathcal{B}(\mathbb{R})$ we have $\lambda(A \cap \mathbb{Q}) \leq \lambda(\mathbb{Q})$, which implies $\lambda(A \cap \mathbb{Q}) = 0$. Thus also, if A has finite measure, $\lambda(A) - \lambda(A \cap \mathbb{I}) = \lambda(A \setminus (A \cap \mathbb{I})) = \lambda(A \cap \mathbb{Q}) = 0$. This is particularly intriguing as it tells us that

$$\lambda(A) = \lambda(A \cap \mathbb{I}),$$

so the only contribution to length of sets of real numbers comes from the irrational numbers. Hence also $\lambda(\mathbb{I}) = \lambda(\mathbb{R} \cap \mathbb{I}) = \lambda(\mathbb{R}) = \infty$ by (L4).

or_zero_meas **Lemma 1.5.5** *The Cantor Set has Lebesgue measure zero.*

PROOF: Recall the construction of the Cantor set $C = \bigcap_{n=1}^{\infty} C_n$ given in Example 1.1.1, and the notation used there. Recall also that the C_n are decreasing, that is $C_{n+1} \subseteq C_n$, and hence also $C \subseteq C_n$ for all n. Since C_n is a union of 2^n disjoint intervals of length 3^{-n} using (M2) and (L3) we have $\lambda(C_n) = 2^n(\frac{1}{3})^n = (\frac{2}{3})^n$. Using monotonicity of measure we thus have $0 \leq \lambda(C) \leq \lambda(C_n) = (\frac{2}{3})^n$. Letting $n \to \infty$, and applying the sandwich rule we obtain $\lambda(C) = 0$. ■

We'll tend to use (L1)-(L5) without explicitly referencing them, from now on. Hopefully, by this point, you're happy to trust that Lebesgue measure matches your intuitive concept of length within $\mathbb{R}$.

1.6 An example of a non-measurable set ($\star$)

non_meas_set

Note that this section has a ($\star$), meaning that it is off-syllabus. It is included for interest.

We might wonder, why go to all the trouble of defining the Borel σ-field? In other words, why can't we measure (the 'size' of) every possible subset of $\mathbb{R}$? We will answer these questions by constructing a strange looking set $\mathscr{V} \subseteq \mathbb{R}$; we will then show that it is not possible to define the Lebesgue measure of $\mathscr{V}$.

As usual, let $\mathbb{Q}$ denote the rational numbers. For any $x \in \mathbb{R}$ we define

$$\mathbb{Q}_x = \{x + q\,;\, q \in \mathbb{Q}\}. \tag{1.7}$$

{eq:Qx_def}

Note that different x values may give the same $\mathbb{Q}_x$. For example, an exercise for you is to prove that $\mathbb{Q}_{\sqrt{2}} = \mathbb{Q}_{1+\sqrt{2}}$. You can think of $\mathbb{Q}_x$ as the set $\mathbb{Q}$ translated by x.

It is easily seen that $\mathbb{Q}_x \cap [0,1]$ is non-empty; just pick some rational q that is slightly less than x and note that $x + (-q) \in \mathbb{Q}_x \cap [0,1]$. Now, for each set $\mathbb{Q}_x$, we pick precisely one element $r \in \mathbb{Q}_x \cap [0,1]$ (it does not matter which element we pick). We write this number r as $r(\mathbb{Q}_x)$. Define

$$\mathscr{V} = \{r(\mathbb{Q}_x)\,;\, x \in \mathbb{R}\},$$

which is a subset of $[0,1]$. For each $q \in \mathbb{Q}$ define

$$\mathscr{V}_q = \{q + m\,;\, m \in \mathscr{V}\}.$$

Clearly $\mathscr{V} = \mathscr{V}_0$, and $\mathscr{V}_q$ is precisely the set $\mathscr{V}$ translated by q. Now, let us record some facts about $\mathscr{V}_q$.

non_meas_pre

Lemma 1.6.1 *It holds that*

1. *If* $q_1 \neq q_2$ *then* $\mathscr{V}_{q_1} \cap \mathscr{V}_{q_2} = \emptyset$.
2. $\mathbb{R} = \bigcup_{q \in \mathbb{Q}} \mathscr{V}_q$.
3. $[0,1] \subseteq \bigcup_{q \in \mathbb{Q} \cap [-1,1]} \mathscr{V}_q \subseteq [-1,2]$.

Before we prove this lemma, let us use it to show that $\mathscr{V}$ cannot have a Lebesgue measure. We will do this by contradiction: assume that $\lambda(\mathscr{V})$ is defined.

Since $\mathscr{V}$ and $\mathscr{V}_q$ are translations of each other, they must have the same Lebesgue measure. We write $c = \lambda(\mathscr{V}) = \lambda(\mathscr{V}_q)$, which does not depend on q. Let us write $\mathbb{Q} \cap [-1,1] = \{q_1, q_2, \ldots,\}$, which we may do because $\mathbb{Q}$ is countable. By parts (1) and (3) of Lemma 1.6.1 and property M(ii) we have

$$\lambda\left(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} \mathscr{V}_q\right) = \sum_{i=1}^{\infty} \lambda(\mathscr{V}_{q_i}) = \sum_{i=1}^{\infty} c.$$

Using the monotonicity property of measures (see Section 1.7) and part (3) of Lemma 1.6.1 we thus have

$$1 \leq \sum_{i=1}^{\infty} c \leq 3.$$

However, there is no value of c which can satisfy this equation! So it is not possible to define of the Lebesgue measure of $\mathscr{V}$. Since we know that we can define the Lebesgue measure on all Borel sets, the set $\mathscr{V}$ is not a Borel set.

The set $\mathscr{V}$ is known as a *Vitali set.* In higher dimensions even stranger things can happen with non-measurable sets; you might like to investigate the *Banach-Tarski paradox.*

Proof: [Of Lemma 1.6.1.] We prove the three claims in turn.

(1) Let $q_1, q_2 \in \mathbb{Q}$ be unequal. Suppose that some $x \in \mathscr{V}_{q_1} \cap \mathscr{V}_{q_2}$ exists – and we now look for a contradiction. By definition of $\mathscr{V}_q$ we have

$$x = q_1 + r(\mathbb{Q}_{x_1}) = q_2 + r(\mathbb{Q}_{x_2}). \tag{1.8}$$ {eq:xqr}

By definition of $\mathbb{Q}_x$ we may write $r(\mathbb{Q}_{x_1}) = x_1 + q_1'$ for some $q_1' \in \mathbb{Q}$, and similarly for x_2, so we obtain $x = q_1 + x_1 + q_1' = q_2 + x_2 + q_2'$ where $q, q' \in \mathbb{Q}$. Hence, setting $q = q_2 - q_1 + q_2' - q_1' \in \mathbb{Q}$, we have $x_1 + q = x_2$, which by (1.7) means that $\mathbb{Q}_{x_1} = \mathbb{Q}_{x_2}$. Thus $r(\mathbb{Q}_{x_1}) = r(\mathbb{Q}_{x_2})$, so going back to (1.8) we obtain that $q_1 = q_2$. But this contradicts our assumption that $q_1 \neq q_2$. Hence x does not exist and $\mathscr{V}_{q_1} \cap \mathscr{V}_{q_2} = \emptyset$.

(2) We will show $\supseteq$ and $\subseteq$. The first is easy: since $\mathscr{V}_q \subseteq \mathbb{R}$ it is immediate that $\mathbb{R} \supseteq \bigcup_{q\in\mathbb{Q}} \mathscr{V}_q$.

Now take some $x \in \mathbb{R}$. Since we may take $q = 0$ in (1.7) we have $x \in \mathbb{Q}_x$. By definition of $r(\mathbb{Q}_x)$ we have $r(\mathbb{Q}_x) = x + q'$ for some $q' \in \mathbb{Q}$. By definition of $\mathscr{V}$ we have $r(\mathbb{Q}_x) \in \mathscr{V}$ and since $x = r(\mathbb{Q}_x) - q'$ we have $x \in \mathscr{V}_{-q'}$. Hence $x \in \bigcup_{q\in\mathbb{Q}} \mathscr{V}_q$.

(3) Since $\mathscr{V} \subseteq [0,1]$, we have $\mathscr{V}_q \cap [0,1] = \emptyset$ whenever $q \notin [-1,1]$. Hence, from part (2) and set algebra we have

$$\mathbb{R} \cap [0,1] \;=\; \left(\bigcup_{q\in\mathbb{Q}} \mathscr{V}_q\right) \cap [0,1] \;=\; \bigcup_{q\in\mathbb{Q}} \mathscr{V}_q \cap [0,1] \;=\; \bigcup_{q\in\mathbb{Q}\cap[-1,1]} \mathscr{V}_q \cap [0,1] \;\subseteq\; \bigcup_{q\in\mathbb{Q}\cap[-1,1]} \mathscr{V}_q.$$

This proves the first $\subseteq$ of (3). For the second simply note that $\mathscr{V} \subseteq [0,1]$ so $\mathscr{V}_q \subseteq [-1,2]$ whenever $q \in [-1,1]$. ■

Remark 1.6.2 *We used the axiom of choice to define the function $r(\cdot)$.*

1.7 Measure and limits

_useful_thms

In this section we return to the consideration of arbitrary measure spaces (S, Σ, m). Let (A_n) be a sequence of sets in Σ. We say that it is *increasing* if $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, and *decreasing* if $A_{n+1} \subseteq A_n$. When (A_n) is increasing, it is easily seen that (A_n^c) is decreasing.

When (A_n) is increasing, a useful technique is the *disjoint union trick* whereby we can write $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$ where the B_ns are all mutually disjoint by defining $B_1 = A_1$ and for $n > 1$, $B_n = A_n - A_{n-1}$. e.g. $\mathbb{R} = \bigcup_{n=1}^\infty [-n, n]$ and here $B_1 = [-1, 1]$, $B_2 = [-2, -1) \cup (1, 2]$ etc.

onotone_meas

Theorem 1.7.1 *Let $A_n \in \Sigma$ for all n. It holds that:*

1. *If (A_n) is increasing and $A = \bigcup_{n=1}^\infty A_n$ then $m(A) = \lim_{n\to\infty} m(A_n)$.*
2. *If (A_n) is decreasing and $A = \bigcap_{n=1}^\infty A_n$, and $m(A_1) < \infty$, then $m(A) = \lim_{n\to\infty} m(A_n)$.*

PROOF: For the second claim, see Problem 1.8. We will prove the first claim here. We use the disjoint union trick and (M2) to find that

$$m(A) = m\left(\bigcup_{n=1}^\infty B_n\right) = \sum_{n=1}^\infty m(B_n) = \lim_{N\to\infty} \sum_{n=1}^N m(B_n) = \lim_{N\to\infty} m\left(\bigcup_{n=1}^N B_n\right) = \lim_{N\to\infty} m(A_N).$$

Here we use that $A_N = B_1 \cup B_2 \cup \cdots \cup B_N$. ■

on_bound_inf

Theorem 1.7.2 (Union bound) *If (A_n) is an arbitrary sequence of sets with $A_n \in \Sigma$ for all $n \in \mathbb{N}$ then*

$$m\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty m(A_n).$$

PROOF: From Problem 1.3, we have $m(A_1 \cup A_2) + m(A_1 \cap A_2) = m(A_1) + m(A_2)$ from which we deduce that $m(A_1 \cup A_2) \le m(A_1) + m(A_2)$. By induction we then obtain for all $N \ge 2$,

$$m\left(\bigcup_{n=1}^N A_n\right) \le \sum_{n=1}^N m(A_n).$$

Now define $X_N = \bigcup_{n=1}^N A_n$. Then $X_N \subseteq X_{N+1}$ and so (X_N) is increasing to $\bigcup_{n=1}^\infty X_n = \bigcup_{n=1}^\infty A_n$. By Theorem 1.7.1 we have

$$m\left(\bigcup_{n=1}^\infty A_n\right) = m\left(\bigcup_{n=1}^\infty X_n\right) = \lim_{N\to\infty} m(X_N) = \lim_{N\to\infty} m\left(\bigcup_{n=1}^N A_n\right) \le \lim_{N\to\infty} \sum_{n=1}^N m(A_n) = \sum_{n=1}^\infty m(A_n).$$

■

1.8 Product Measures

prod_meas_350

We calculate areas of rectangles by multiplying products of lengths of their sides. This suggests trying to formulate a theory of products of measures. Let (S_1, Σ_1, m_1) and (S_2, Σ_2, m_2) be two measure spaces. Form the Cartesian product $S_1 \times S_2$. We can similarly try to form a product of σ-fields

$$\Sigma_1 \times \Sigma_2 = \{A \times B; A \in \Sigma_1, B \in \Sigma_2\},$$

but it turns out that $\Sigma_1 \times \Sigma_2$ is not a σ-field in general e.g. take $\Sigma_1 = \Sigma_2 = \mathbb{R}$ and note that $((0,1) \times (0,1))^c$ is not a rectangle. Instead the object we want is $\Sigma_1 \otimes \Sigma_2$, which is defined to be the smallest σ-field containing all the sets in $\Sigma_1 \times \Sigma_2$. We state but do not prove:

Theorem 1.8.1 *There exists a measure $m_1 \times m_2$ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ such that*

$$(m_1 \times m_2)(A \times B) = m_1(A)m_2(B) \tag{1.9}$$

{eq:prod_mea

for all $A \in \Sigma_1, B \in \Sigma_2$.

Definition 1.8.2 The measure $m_1 \times m_2$ is called the product measure of m_1 and m_2.

For example, consider $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. We equip it with the Borel σ-field, $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Then the product Lebesgue measure $\lambda_2 = \lambda \times \lambda$ has the property that

$$\lambda_2((a,b) \times (c,d)) = (b-a)(d-c).$$

Of course, $(b-a)(d-c)$ is the area of the rectangle $(a,b) \times (c,d)$. In fact, from a mathematical point of view the measure λ_2 is the *definition* of area. Similarly, $\lambda_3 = \lambda \times \lambda \times \lambda$ is how we define volume, in three dimensions.

Remark 1.8.3 *After thinking about $\lambda \times \lambda \times \lambda$, we might ask if, given measures m_1, m_2, m_3, we have $(m_1 \times m_2) \times m_3 = m_1 \times (m_2 \times m_3)$. It is true, but we won't prove it. Consequently we write both these as simply $m_1 \times m_2 \times m_3$, without any ambiguity.*

We can go beyond 3 dimensions. Given n-measure spaces $(S_1, \Sigma_1, m_1), (S_2, \Sigma_2, m_2), \ldots, (S_n, \Sigma_n, m_n)$, we can iterate the above procedure to define the product σ-field $\Sigma_1 \otimes \Sigma_2 \otimes \cdots \otimes \Sigma_n$ and the product measure $m_1 \times m_2 \times \cdots \times m_n$ so that for $A_i \in \Sigma_i, 1 \leq i \leq n$,

$$(m_1 \times m_2 \times \cdots \times m_n)(A_1 \times A_2 \times \cdots \times A_n) = m_1(A_1)m_2(A_2) \cdots m_n(A_n).$$

In particular n-dimensional Lebesgue measure on $\mathbb{R}^n$ may be defined in this way.

Of course there are many measures that one can construct on $(S_1 \times S_2, \Sigma_1 \times \Sigma_2)$ and not all of these will be product measures. In probability spaces, product measures are closely related to the notion of independence, as we will see later. If you write $m_1 = m_2 = \mathbb{P}$ in (1.9) you might be able to see why.

1.9 Exercises

ps:size_Pn

1.1 Show that if S is a set containing n elements, then the power set $\mathcal{P}(S)$ contains 2^n elements.

Hint: How many subsets are there of size r, for a fixed $1 \le r \le n$? The binomial theorem may also be of some use.

:sigma_union

1.2 Let Σ_1 and Σ_2 be σ-fields of subsets of a set S. Define

$$\Sigma_1 \cap \Sigma_2 = \{A \subseteq S; A \in \Sigma_1 \text{ and } A \in \Sigma_2\}.$$

Show that $\Sigma_1 \cap \Sigma_2$ is a σ-field. Define $\Sigma_1 \cup \Sigma_2 = \{A \subseteq S; A \in \Sigma_1 \text{ or } A \in \Sigma_2\}$. Why is $\Sigma_1 \cup \Sigma_2$ not in general a σ-field?

easure_basic

1.3 If (S, Σ, m) is a measure space, show that for all $A, B \in \Sigma$

(a) $m(A \cup B) + m(A \cap B) = m(A) + m(B)$,

(b) $m(A \cup B) \le m(A) + m(B)$.

Hence prove that if $A_1, A_2, \ldots, A_n \in \Sigma$,

$$m\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} m(A_i).$$

form_measure

1.4 (a) If m is a measure on (S, Σ) and $k > 0$, show that km is also a measure on (S, Σ) where for all $A \in \Sigma$,

$$(km)(A) = km(A).$$

Hence show that if m is a finite measure and $m(S) \ne 0$, then $\mathbb{P}$ is a probability measure where $\mathbb{P}(A) = \frac{m(A)}{m(S)}$ for all $A \in \Sigma$.

(b) Let $[a, b]$ be a finite closed interval in $\mathbb{R}$. Write down a formula for the *uniform distribution* as a probability measure on $([a, b], \mathcal{B}([a, b]))$, using the above considerations and Lebesgue measure.

Hint: Recall that the uniform distribution has the property that subintervals of $[a, b]$ which have the same length, will have the same probability.

(c) If m and n are measures on (S, Σ), deduce that $m + n$ is a measure on (S, Σ) where $(m + n)(A) = m(A) + n(A)$ for all $A \in \Sigma$.

onal_measure

1.5 (a) If m is a measure on (S, Σ) and $B \in \Sigma$ is fixed, show that $m_B(A) = m(A \cap B)$ for $A \in \Sigma$ defines another measure on (S, Σ).

(b) If m is a finite measure and $m(B) > 0$, deduce that $\mathbb{P}_B$ is a probability measure where

$$\mathbb{P}_B(A) = \frac{m_B(A)}{m(B)}.$$

How does this relate to the notion of conditional probability?

prob_measure

1.6 Let $S = \{x_1, x_2, \ldots, x_n\}$ be a finite set and let $c_1, c_2, \ldots, c_n$ be non-negative numbers. Deduce that m is a measure on $(S, \mathcal{P}(S))$ where $m = \sum_{i=1}^{n} c_i \delta_{x_i}$. What condition should be imposed on $\{c_1, c_2, \ldots, c_n\}$ for m to be a probability measure?

borel_closed **1.7** Show that $\mathcal{B}(\mathbb{R})$ contains all closed intervals $[a,b]$, where $-\infty < a < b < \infty$.

s_decreasing **1.8** (a) Let m be a finite measure on (S,Σ).

(i) Show that, for any $A \in \Sigma$,

$$m(A^c) = m(S) - m(A).$$

(ii) Let $(A_n)_{n\in\mathbb{N}}$ be a decreasing sequence of sets in Σ. Show that $m(A_n) \to m(\cap_j A_j)$ as $n \to \infty$.

(b) Give an example of a (not finite!) measure m, and sets A_n and $A = \bigcap_{n=1}^{\infty} A_n$ such that $m(A) \neq \lim_{n\to\infty} m(A_n)$.

Challenge Questions

_sigma_field **1.9** Let S be a finite set and Σ be a σ-field on S. Consider the set

$$\Pi = \{A \in \Sigma\,;\text{ if } B \in \Sigma \text{ and } B \subseteq A \text{ then either } B = A \text{ or } B = \emptyset\}. \qquad (\star)$$ {eq:atoms}

(a) Show that Π is a finite set.

(b) Using (a), let us enumerate the elements of Π as $\Pi = \{\Pi_1, \Pi_2, \cdots, \Pi_k\}$, where each Π_i is distinct from the others.

(i) Show that $\Pi_i \cap \Pi_j = \emptyset$ for $i \neq j$. *Hint: Could $\Pi_i \cap \Pi_j$ be an element of Π?*

(ii) Show that $\cup_{i=1}^k \Pi_i = S$. *Hint: If $C = S \setminus \cup_{i=1}^k \Pi_i$ is non-empty, is $C \in \Pi$?*

(iii) Let $A \in \Sigma$. Show that

$$A = \bigcup_{i\in I} \Pi_i$$

where $I = \{i = 1,\ldots,k\,;\ A \cap \Pi_i \neq \emptyset\}$.

cantor_false **1.10** Prove that both of the follow claims are false.

(a) The Cantor set C contains an open interval $(a,b) \subseteq C$, where $a < b$.

(b) If a Borel set has non-zero Lebesgue measure then it contains an open interval.

Chapter 2

Measurable Functions

urable_funcs

We now restrict ourselves to studying a particular kind of function, known as a *measurable* function. For measure theory, this is an important step, because it allows us to exclude some very strangely behaved examples (in the style of Section 1.6) that would disrupt our theory.

We will see in Chapter 4 that measurable functions also play a huge role in probability theory, where they provide a mechanism for identifying which random variables depend on which information.

2.1 Liminf and Limsup

This section introduces some important tools from analysis which there wasn't time to cover in MAS221. Let (a_n) be a sequence of real numbers. It may or may not converge. For example the sequence whose nth term is $(-1)^n$ fails to converge but it does have two convergent subsequences corresponding to $a_{2n-1} = -1$ and $a_{2n} = 1$. This is a very special case of a general phenomenon that we'll now describe.

Assume that the sequence (a_n) is *bounded*, i.e. there exists $K > 0$ so that $|a_n| \leq K$ for all $n \in \mathbb{N}$. Define a new sequence (b_n) by $b_n = \inf_{k \geq n} a_k$. Then you can check that (b_n) is monotonic increasing and bounded above (by K). Hence it converges to a limit.

In fact, we are fine if (a_n) is not bounded. In this case b_n is still monotone increasing, but we now have the extra possibilities that $b_n \to \infty$ or $b_n \to -\infty$. In fact, we might even have $b_n = \pm\infty$ for some or all finite n.

We define $\liminf_{n\to\infty} a_n = \lim_{n\to\infty} b_n$, where $b_n = \inf_{k\geq n} a_k$. We call this the *limit inferior* or *liminf* of the sequence (a_n). So we have

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf_{k\geq n} a_k = \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k. \tag{2.1}$$ {li}

Similarly the sequence (c_n) where $c_n = \sup_{k\geq n} a_k$ is monotonic decreasing and bounded below. So it also converges to a limit which we call the *limit superior* or *limsup* for short. We denote $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} c_n$. Then we have

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k = \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k. \tag{2.2}$$ {prop:ls}

Clearly we have $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$. In Problem **2.2**, you can investigate some other properties of $\limsup$ and $\liminf$. It can be shown that the smallest limit of any convergent subsequence of (a_n) is $\liminf_{n\to\infty} a_n$, and the largest limit is $\limsup_{n\to\infty} a_n$. The next theorem is very useful:

thm:lils **Theorem 2.1.1** *A bounded sequence of real numbers (a_n) converges to a limit if and only if $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$. In this case we have*

$$\lim_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n.$$

PROOF: If (a_n) converges to a limit, then all of its subsequences also converge to the same limit and it follows that $\lim_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$. Conversely suppose that we don't know that (a_n) converges but we do know that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$. Then for all $n \in \mathbb{N}$,

$$0 \le a_n - \inf_{k\ge n} a_k \le \sup_{k\ge n} a_k - \inf_{k\ge n} a_k.$$

But

$$\lim_{n\to\infty}\left(\sup_{k\ge n} a_k - \inf_{k\ge n} a_k\right) = \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} a_n = 0$$

and so

$$\lim_{n\to\infty}\left(a_n - \inf_{k\ge n} a_k\right) = 0$$

by the sandwich rule. But since

$$a_n = \left(a_n - \inf_{k\ge n} a_k\right) + \inf_{k\ge n} a_k,$$

and $\lim_{n\to\infty} \inf_{k\ge n} a_k = \liminf_{n\to\infty} a_n$, we can use the algebra of limits to deduce that (a_n) converges to the common value of $\liminf_{n\to\infty} a_n$ and $\limsup_{n\to\infty} a_n$. ■

2.2 Measurable Functions - Basic Concepts

eas_funcs_def

In Section 1.6 we saw the existence subsets of $\mathbb{R}$ that were not measurable, with respect to Lebesgue measure. These sets had no meaningful concept of 'length'. We are looking to build a theory of integration in Chapter 3, but for such A integrating the indicator function $\mathbb{1}_A$ would be dangerously close to trying to measure the length of A, which we know we cannot do. Since $\mathbb{1}_A$ is a function, we conclude that, just as with sets, we need a similar concept of a measurable *function*. In this section, we deal with functions $f : S \to \mathbb{R}$, where (S, Σ) is a measurable space.

We say that a function $f : S \to \mathbb{R}$ is *measurable* if $f^{-1}((a, \infty)) \in \Sigma$ for all $a \in \mathbb{R}$. We will spend the rest of this section finding lots of different, and mathematically equivalent, ways of writing this same definition.

eas_halfints

Theorem 2.2.1 *Let $f : S \to \mathbb{R}$. The following are equivalent:*

(i) $f^{-1}((a, \infty)) \in \Sigma$ for all $a \in \mathbb{R}$ (i.e. f is measurable).

(ii) $f^{-1}([a, \infty)) \in \Sigma$ for all $a \in \mathbb{R}$.

(iii) $f^{-1}((-\infty, a)) \in \Sigma$ for all $a \in \mathbb{R}$.

(iv) $f^{-1}((-\infty, a]) \in \Sigma$ for all $a \in \mathbb{R}$.

Proof: (i) $\Leftrightarrow$ (iv) as $f^{-1}(A)^c = f^{-1}(A^c)$ and Σ is closed under taking complements.

(ii) $\Leftrightarrow$ (iii) is proved similarly.

(i) $\Rightarrow$ (ii) uses $[a, \infty) = \bigcap_{n=1}^{\infty}(a - 1/n, \infty)$ and so

$$f^{-1}([a, \infty)) = \bigcap_{n=1}^{\infty} f^{-1}((a - 1/n, \infty))$$

and the result follows since Σ is closed under countable intersections.

(ii) $\Rightarrow$ (i) uses

$$f^{-1}((a, \infty)) = \bigcup_{n=1}^{\infty} f^{-1}([a + 1/n, \infty))$$

and the fact that Σ is closed under countable unions. ■

It follows that f is measurable if any of (i) to (iv) in Theorem 2.2.1 is established for all $a \in \mathbb{R}$. In Problem **2.4** you can show that f is measurable if and only if $f^{-1}((a, b)) \in \Sigma$ for all $-\infty \le a < b \le \infty$. With similar ideas we can also use closed and half-open intervals.

A set O in $\mathbb{R}$ is *open* if for every $x \in O$ there is an open interval I containing x for which $I \subseteq O$. It follows that every open interval in $\mathbb{R}$ is an open set. We might ask what other kinds of open set there are. The following result gives a surprisingly clear answer, which implies that all open subsets of $\mathbb{R}$ are Borel sets.

nt_intervals

Proposition 2.2.2 *Every open set O in $\mathbb{R}$ is a countable union of disjoint open intervals.*

Proof: Note that a 'countable union' includes the case where we only need finitely many intervals. Let us first note that if O_i are opens sets for all $i \in I$ then (even if I is uncountable) the set $O = \cup_i O_i$ is open. See Exercise **2.12** for a proof of this fact.

For $x \in O$, let I_x be the union of all open intervals containing x for which $I_x \subseteq O$. Then I_x is open. Also, I_x is an interval, because if $a < b < c$ with $a, b \in I_x$ then there are open intervals $(a - \epsilon_1, x + \epsilon_2)$ and $(x - \epsilon_3, c + \epsilon_4)$ within I_x and b is within their union, so $b \in I_x$.

If $x, y \in O$ and $x \neq y$ then I_x and I_y are either disjoint or identical. To see this, note that if $I_x \cap I_y$ is non-empty then $I_x \cup I_y$ is a non-empty open interval contained within O, which implies $I_x \cup I_y$ is also contained within both I_x and I_y. Thus $I_x = I_y$.

However, there can only be countably many different I_x, because we can only fit at most countably many (non-empty) disjoint open intervals within $\mathbb{R}$. We now select a rational number $r(x)$ in every distinct I_x and rewrite O as the countable disjoint union over intervals I_x labelled by distinct rationals $r(x)$. ■

en_set_borel **Remark 2.2.3** *By Proposition 2.2.2, every open subset of $\mathbb{R}$ is a Borel set.*

hm:meas_open **Theorem 2.2.4** *The mapping $f : S \to \mathbb{R}$ is measurable if and only if $f^{-1}(O) \in \Sigma$ for all open sets O in $\mathbb{R}$.*

Proof: Suppose that $f^{-1}(O) \in \Sigma$ for all open sets O in $\mathbb{R}$. Then in particular $f^{-1}((a, \infty)) \in \Sigma$ for all $a \in \mathbb{R}$ and so f is measurable. Conversely assume that O is open in $\mathbb{R}$ and use Proposition 2.2.2 to write $O = \bigcup_{n=1}^{\infty}(a_n, b_n)$. Then

$$f^{-1}(O) = \bigcup_{n=1}^{\infty} f^{-1}((a_n, b_n)).$$

If f is measurable, then $f^{-1}((a_n, b_n)) \in \Sigma$ for all $n \in \mathbb{N}$ by Problem 2.4, and so $f^{-1}(O) \in \Sigma$ since Σ is closed under countable unions. ■

m:meas_borel **Theorem 2.2.5** *The mapping $f : S \to \mathbb{R}$ is measurable if and only if $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{B}(\mathbb{R})$.*

Proof: Suppose that f is measurable and let $\mathcal{A} = \{E \subseteq \mathbb{R}; f^{-1}(E) \in \Sigma\}$. We first show that $\mathcal{A}$ is a σ-field.

(S1). $\mathbb{R} \in \mathcal{A}$ as $S = f^{-1}(\mathbb{R})$.

(S2). If $E \in \mathcal{A}$ then $E^c \in \mathcal{A}$ since $f^{-1}(E^c) = f^{-1}(E)^c \in \Sigma$.

(S3). If (A_n) is a sequence of sets in $\mathcal{A}$ then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ since $f^{-1}\left(\bigcup_n A_n\right) = \bigcup_n f^{-1}(A_n) \in \Sigma$.

By Problem 2.4, $f^{-1}((a, b)) \in \Sigma$ for all $-\infty \leq a < b \leq \infty$, and so $\mathcal{A}$ is a σ-field of subsets of $\mathbb{R}$ that contains all the open intervals. But by definition, $\mathcal{B}(\mathbb{R})$ is the smallest such σ-field. It follows that $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{A}$ and so $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{B}(\mathbb{R})$.

The converse is easy (e.g. just allow A to range over open sets, and use Theorem 2.2.4). ■

Remark 2.2.6 *($\star$) Based on Theorem 2.2.5, the idea of a measurable function can be extended beyond the real numbers. Let (S_1, Σ_1) and (S_2, Σ_2) be measurable spaces. The mapping $f : S_1 \to S_2$ is said to be measurable if $f^{-1}(A) \in \Sigma_1$ for all $A \in \Sigma_2$. We'll stay within the real numbers in this course.*

2.3 Examples of Measurable Functions

We first consider the case where $S = \mathbb{R}$ (equipped with its Borel σ-field) and look for classes of measurable functions. In fact we will prove that

$$\{\text{continuous functions on } \mathbb{R}\} \subseteq \{\text{measurable functions on } \mathbb{R}\}.$$

First we present a result that is well-known (in the wider context of continuous functions on metric spaces) to those who have taken MAS331.

prop:cts_open **Proposition 2.3.1** *A mapping $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if $f^{-1}(O)$ is open for every open set O in $\mathbb{R}$.*

PROOF: First suppose that f is continuous. Choose an open set O and let $a \in f^{-1}(O)$ so that $f(a) \in O$. Then there exists $\epsilon > 0$ so that $(f(a) - \epsilon, f(a) + \epsilon) \subseteq O$. By definition of continuity of f, for such an ϵ there exists $\delta > 0$ so that $x \in (a - \delta, a + \delta) \Rightarrow f(x) \in (f(a) - \epsilon, f(a) + \epsilon)$. But this tells us that $(a - \delta, a + \delta) \subseteq f^{-1}((f(a) - \epsilon, f(a) + \epsilon)) \subseteq f^{-1}(O)$. Since a is arbitrary we conclude that $f^{-1}(O)$ is open.

Conversely, suppose that $f^{-1}(O)$ is open for every open set O in $\mathbb{R}$. Choose $a \in \mathbb{R}$ and let $\epsilon > 0$. Then since $(f(a)-\epsilon, f(a)+\epsilon)$ is open so is $f^{-1}((f(a)-\epsilon, f(a)+\epsilon))$. Since $a \in f^{-1}((f(a)-\epsilon, f(a)+\epsilon))$ there exists $\delta > 0$ so that $(a - \delta, a + \delta) \subseteq f^{-1}((f(a) - \epsilon, f(a) + \epsilon))$. From here you can see that whenever $|x - a| < \delta$ we must have $|f(x) - f(a)| < \epsilon$. But then f is continuous at a and the result follows. ■

cor:meas_cts **Corollary 2.3.2** *Every continuous function on $\mathbb{R}$ is measurable.*

PROOF: Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and O be an arbitrary open set in $\mathbb{R}$. Then by Proposition 2.3.1, $f^{-1}(O)$ is an open set in $\mathbb{R}$. Then $f^{-1}(O)$ is in $\mathcal{B}(\mathbb{R})$ by Proposition 2.2.2 and Remark 2.2.3. Hence f is measurable by Theorem 2.2.4. ■

There are many discontinuous functions on $\mathbb{R}$ that are also measurable. Lets look at an important class of examples in a wider context. Let (S, Σ) be a general measurable space. Fix $A \in \Sigma$ and define the *indicator function* $\mathbb{1}_A : S \to \mathbb{R}$ by

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

To see that it is measurable its enough to check that

$$\begin{aligned} \mathbb{1}_A^{-1}((c, \infty)) &= \emptyset \in \Sigma \qquad \text{if } c \geq 1 \\ \mathbb{1}_A^{-1}((c, \infty)) &= A \in \Sigma \qquad \text{if } 0 \leq c < 1 \\ \mathbb{1}_A^{-1}((c, \infty)) &= S \in \Sigma \qquad \text{if } c < 0 \end{aligned}$$

If $(S, \Sigma) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ or indeed if S is any metric space, then $\mathbb{1}_A$ is clearly a measurable but discontinuous function.

A particularly interesting example is obtained by taking $(S, \Sigma) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $A = \mathbb{Q}$. Then $\mathbb{1}_A$ is called *Dirichlet's jump function*. We have already seen that $\mathbb{Q}$ is measurable (it is a countable union of points). As there is a rational number between any pair of irrationals and an irrational number between any pair of rationals, we see that in this case $\mathbb{1}_A$ is measurable, but discontinuous at every point of $\mathbb{R}$.

A measurable function from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is sometimes called *Borel measurable*.

2.4 Algebra of Measurable Functions

Our main goals in this section is to show that, sums, products, limits etc of measurable functions are themselves measurable. Consequently, almost anything we can think of doing with measurable functions will just give us back more measurable functions – just like the situation we already established for measurable sets. Throughout this section, (S, Σ) is a measurable space.

We now introduce several pieces of notation. Let f and g be functions from S to $\mathbb{R}$ and define for all $x \in S$,

$$(f \vee g)(x) = \max\{f(x), g(x)\}, \qquad (f \wedge g)(x) = \min\{f(x), g(x)\}.$$

This is called *pointwise* definition, because we define e.g. $f \vee g$ by evaluating each of the individual functions f and g at the point x. Similarly, given functions f and g we define new functions $f+g$, fg and $-f$ by $(f+g)(x) = f(g) + g(x)$, $(fg)(x) = f(x)g(x)$, $(-f)(x) = -f(x)$. For $\alpha \in \mathbb{R}$ we define $(\alpha f)(x) = \alpha f(x)$.

Let $\mathbf{0}$ denote the zero function that maps every element of S to zero, i.e. $\mathbf{0} = \mathbb{1}_\emptyset$. Then $\mathbf{0}$ is measurable since it is the indicator factor of a measurable set (or use Problem **2.3**). Define $f_+ = f \vee \mathbf{0}$ and $f_- = -f \vee \mathbf{0}$. More explicitly,

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}, \qquad f_-(x) = \begin{cases} 0 & \text{if } f(x) \geq 0 \\ -f(x) & \text{if } f(x) < 0 \end{cases}.$$

We now consider measurability of all these functions.

_min_max_etc **Theorem 2.4.1** *Let f, g be measurable functions from $S \to \mathbb{R}$. Then:*

1. *$f \vee g$ and $f \wedge g$ are measurable.*
2. *αf is measurable, for any $\alpha \in \mathbb{R}$.*
3. *f_+ and f_- are measurable.*

PROOF: For part 1, note that for all $c \in \mathbb{R}$ we have

$$\begin{aligned}(f \vee g)^{-1}((c,\infty)) &= f^{-1}((c,\infty)) \cup g^{-1}((c,\infty)) \\ (f \wedge g)^{-1}((c,\infty)) &= f^{-1}((c,\infty)) \cap g^{-1}((c,\infty)).\end{aligned}$$

We have that $f^{-1}((c,\infty))$ is measurable, hence the right hand side is also measurable. By Theorem 2.2.1, $f \vee g$ and $f \wedge g$ are measurable. Part 2 is Problem **2.5**. Taking $\alpha = -1$ in part 2 gives that $-f$ is measurable. Given this, the rest of part 2 follows immediately from part 1. ∎

We write $\{f > g\} = \{x \in S\,;\, f(x) > g(x)\}$. We'll use similar notation for multiple inequalities of all types, and for constants. For example if $a, b \in \mathbb{R}$ then $\{a \leq f < c\} = \{x \in S\,;\, a \leq f(x) < c\}$.

_f_greater_g **Lemma 2.4.2** *If f and g are measurable then $\{f > g\} \in \Sigma$.*

PROOF: Let $\{r_n, n \in \mathbb{N}\}$ be an enumeration of the rational numbers. Then

$$\begin{aligned}\{f > g\} &= \bigcup_{n\in\mathbb{N}} \{f > r_n > g\} \\ &= \bigcup_{n\in\mathbb{N}} \{f > r_n\} \cap \{g < r_n\} \\ &= \bigcup_{n\in\mathbb{N}} f^{-1}((r_n,\infty)) \cap g^{-1}((-\infty, r_n)) \in \Sigma\end{aligned}$$

■

eas_plus_etc **Theorem 2.4.3** *Let f, g be measurable functions from $S \to \mathbb{R}$. Then:*

1. *$f + g$ and fg are measurable.*
2. *If $G : \mathbb{R} \to \mathbb{R}$ is continuous then $G \circ f$ is measurable, as a function from S to $\mathbb{R}$.*

PROOF: Let us first show that $f + g$ is measurable. From Problem **2.5**, we have that $a - g$ is measurable for all $a \in \mathbb{R}$. Now

$$(f + g)^{-1}((a, \infty)) = \{f + g > a\} = \{f > a - g\} \in \Sigma,$$

by Lemma 2.4.2 and this establishes the result.

Next, we'll prove all of part 2. We have $(G \circ f)(x) = G(f(x))$ so for any $A \subseteq \mathbb{R}$ we have $(G \circ f)^{-1}(A) = f^{-1}(G^{-1}(A))$. Let $A \in \mathcal{B}(\mathbb{R})$. Since G is continuous, $G^{-1}(A) \in \mathcal{B}(\mathbb{R})$ by Corollary 2.3.2 and Theorem 2.2.5. Hence $f^{-1}(G^{-1}(A)) \in \Sigma$ by applying Theorem 2.2.5 to f. Since A was an arbitrary Borel set, the function $(G \circ f) : S \to \mathbb{R}$ is measurable by Theorem 2.2.5.

It remains to show that fg is measurable. Note that

$$fg = \frac{1}{4}[(f + g)^2 - (f - g)^2].$$

Using part 2 with $G(x) = x^2$ we deduce that h^2 is measurable whenever h is, and we have already shown that sums of measurable functions are measurable. The result follows. ■

We can use induction to show that if $f_1, f_2, \ldots, f_n$ are measurable and $c_1, c_2, \ldots, c_n \in \mathbb{R}$ then f is also measurable where $f = c_1 f_1 + c_2 f_2 + \cdots + c_n f_n$. So the set of measurable functions from S to $\mathbb{R}$ forms a real vector space. Of particular interest are the *simple functions* which take the form $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ where $A_i \in \Sigma$ $(1 \leq i \leq n)$. We will learn more about this highly useful class of functions in Section 2.5 and in Chapter 3 we will see that they play an important role in integration theory.

2.4.1 Limits of Measurable Functions

Let (f_n) be a bounded[1] sequence of functions from S to $\mathbb{R}$ such that we also have the condition $\sup_{n\in\mathbb{N}}\sup_{x\in S}|f_n(x)|<\infty$. Define $\inf_{n\in\mathbb{N}} f_n$ and $\sup_{n\in\mathbb{N}} f_n$ pointwise, that is

$$\left(\inf_{n\in\mathbb{N}} f_n\right)(x)=\inf_{n\in\mathbb{N}} f_n(x) \quad \text{and} \quad \left(\sup_{n\in\mathbb{N}} f_n\right)(x)=\sup_{n\in\mathbb{N}} f_n(x)$$

for all $x\in S$. Similarly, define $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ pointwise,

$$\left(\liminf_{n\to\infty} f_n\right)(x)=\liminf_{n\to\infty} f_n(x) \quad \text{and} \quad \left(\limsup_{n\to\infty} f_n\right)(x)=\limsup_{n\to\infty} f_n(x)$$

for all $x\in S$. Lastly, we say that the sequence (f_n) converges *pointwise* to f as $n\to\infty$ if $\lim_{n\to\infty} f_n(x)=f(x)$ for all $x\in S$.

thm:meas_lim

Theorem 2.4.4 *Let (f_n) be a sequence of measurable functions. Then:*

1. $\inf_{n\in\mathbb{N}} f_n$ *and* $\sup_{n\in\mathbb{N}} f_n$ *are measurable.*
2. $\liminf_{n\to\infty} f_n$ *and* $\limsup_{n\to\infty} f_n$ *are both measurable.*
3. *If (f_n) converges pointwise to f as $n\to\infty$, then f is measurable.*

PROOF: For all $c\in\mathbb{R}$,

$$\left(\inf_{n\in\mathbb{N}} f_n\right)^{-1}([c,\infty))=\left\{\inf_{n\in\mathbb{N}} f_n\geq c\right\}=\bigcap_{n\in\mathbb{N}}\{f_n\geq c\}=\bigcap_{n\in\mathbb{N}} f^{-1}([c,\infty)),$$
$$\left(\sup_{n\in\mathbb{N}} f_n\right)^{-1}((c,\infty))=\left\{\sup_{n\in\mathbb{N}} f_n> c\right\}=\bigcup_{n\in\mathbb{N}}\{f_n> c\}=\bigcup_{n\in\mathbb{N}} f^{-1}((c,\infty)).$$

In both cases the right hand is in Σ, hence so is the left hand side. Thus $\inf_n f_n$ and $\sup_n f_n$ are measurable.

For part 2, recall that $\liminf_n f_n=\sup_n\inf_{k\geq n} f_k$ and $\limsup_n f_n=\inf_n\sup_{k\geq n} f_k$. By several applications of part 1, we have that $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ are measurable. For part 3, by Theorem 2.1.1 $f(x)=\liminf_{n\to\infty} f_n(x)$ for all $x\in S$ and so f is measurable by part 2. ■

[1] We can drop the boundedness requirement if we work with functions taking values in $[-\infty,\infty]$. See Section 2.6.

2.5 Simple Functions

simple_funcs

Recall the definition of indicator functions $\mathbb{1}_A$ where $A \in \Sigma$. A mapping $f : S \to \mathbb{R}$ is said to be *simple* if it takes the form

$$f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i} \tag{2.3}$$ {simple}

where $c_1, c_2, \ldots, c_n \in \mathbb{R}$ and $A_1, A_2, \ldots, A_n \in \Sigma$ with $\bigcup_{i=1}^n A_i = S$ and $A_i \cap A_j = \emptyset$ when $i \neq j$. In other words a simple function is a (finite) linear combination of indicator functions of non-overlapping sets. It follows from Theorems 2.4.1 and 2.4.3 that every simple function is measurable. It is straightforward to prove that sums and scalar multiples of simple functions are themselves simple, so the set of all simple functions form a vector space.

We now prove a key result that shows that simple functions are powerful tools for approximating measurable functions. Recall that a mapping $f : S \to \mathbb{R}$ is *non-negative* if $f(x) \geq 0$ for all $x \in S$, which we write for short as $f \geq 0$. We write $f \leq g$ when $g - f \geq 0$. It is easy to see that a simple function of the form (2.3) (with $A_i \neq \emptyset$ for all $i = 1, \ldots, n$) is non-negative if and only if $c_i \geq 0$ $(1 \leq i \leq n)$.

simple_approx **Theorem 2.5.1** *Let $f : S \to \mathbb{R}$ be measurable and non-negative. Then there exists a sequence (s_n) of non-negative simple functions on S with $s_n \leq s_{n+1} \leq f$ for all $n \in \mathbb{N}$ so that (s_n) converges pointwise to f as $n \to \infty$. Moreover, if f is bounded then the convergence is uniform.*

PROOF: We split this into three parts.

Step 1 Construction of (s_n).

Divide the interval $[0, n)$ into $n2^n$ subintervals $\{I_j, 1 \leq j \leq n2^n\}$, each of length $\frac{1}{2^n}$ by taking $I_j = \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right)$. Let $E_j = f^{-1}(I_j)$ and $F_n = f^{-1}([n, \infty))$. Then $S = \bigcup_{j=1}^{n2^n} E_j \cup F_n$. We define for all $x \in S$

$$s_n(x) = \sum_{j=1}^{n2^n} \left(\frac{j-1}{2^n}\right) \mathbb{1}_{E_j}(x) + n\mathbb{1}_{F_n}(x).$$

Step 2 Properties of (s_n).

For $x \in E_j, s_n(x) = \frac{j-1}{2^n}$ and $\frac{j-1}{2^n} \leq f(x) < \frac{j}{2^n}$ and so $s_n(x) \leq f(x)$. For $x \in F_n, s_n(x) = n$ and $f(x) \geq n$. So we conclude that $s_n \leq f$ for all $n \in \mathbb{N}$.

To show that $s_n \leq s_{n+1}$, fix an arbitrary j and consider $I_j = \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right)$. For convenience, we write I_j as I and we observe that $I = I_1 \cup I_2$ where $I_1 = \left[\frac{2j-2}{2^{n+1}}, \frac{2j-1}{2^{n+1}}\right)$ and $I_2 = \left[\frac{2j-1}{2^{n+1}}, \frac{2j}{2^{n+1}}\right)$. Let $E = f^{-1}(I), E_1 = f^{-1}(I_1)$ and $E_2 = f^{-1}(I_2)$. Then $s_n(x) = \frac{j-1}{2^n}$ for all $x \in E, s_{n+1}(x) = \frac{j-1}{2^n}$ for all $x \in E_1$, and $s_{n+1}(x) = \frac{2j-1}{2^{n+1}}$ for all $x \in E_2$. It follows that $s_n \leq s_{n+1}$ for all $x \in E$. A similar (easier) argument can be used on F_n.

Step 3 Convergence of (s_n).

Fix any $x \in S$. Since $f(x) \in \mathbb{R}$ there exists $n_0 \in \mathbb{N}$ so that $f(x) \leq n_0$. Then for each $n > n_0, f(x) \in I_j$ for some $1 \leq j \leq n2^n$, i.e. $\frac{j-1}{2^n} \leq f(x) < \frac{j}{2^n}$. But $s_n(x) = \frac{j-1}{2^n}$ and so $|f(x) - s_n(x)| < \frac{1}{2^n}$ and the result follows. If f is bounded we can find $n_0 \in \mathbb{N}$ so that $f(x) \leq n_0$ for all $x \in \mathbb{R}$. Then the argument just given yields $|f(x) - s_n(x)| < \frac{1}{2^n}$ for all $x \in \mathbb{R}$ from which we can deduce the uniformity of the convergence. ■

2.6 Extended Real Functions

sec:ext_real

Let (S, Σ, m) be a measure space. An *extended function* on S is a mapping $f : S \to \mathbb{R}^*$ where $\mathbb{R}^* = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ is the extended real number line. Theorem 2.2.1 extends easily to this context and we have $f^{-1}((a, \infty]) \in \Sigma$ for all $a \in \mathbb{R}$, if and only if $f^{-1}([a, \infty]) \in \Sigma$ for all $a \in \mathbb{R}$, if and only if $f^{-1}([-\infty, a)) \in \Sigma$ for all $a \in \mathbb{R}$, if and only if $f^{-1}([-\infty, a]) \in \Sigma$ for all $a \in \mathbb{R}$. We then say that f is measurable if it satisfies any one (and hence all) of these conditions. Now suppose that (f_n) is a sequence of measurable functions from S to $[0, \infty)$. If the functions are not bounded, then there may exist a set $A \in \Sigma$ with $m(A) > 0$ so that $\liminf_{n\to\infty} f_n(x) = \infty$ for all $x \in A$. Then we may regard $\liminf_{n\to\infty} f_n$ as an extended measurable function in the sense just given. This will be used implicitly in the integration theory that we'll describe in the next chapter.

2.7 Exercises

icator_funcs **2.1** Let (S, Σ) be a measurable space. Show that for all $A, B \in \Sigma$

(a) $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}$,

(b) $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$,

(c) $\mathbb{1}_{A-B} = \mathbb{1}_A - \mathbb{1}_B$, if $B \subseteq A$,

(d) $\mathbb{1}_{A\cap B} = \mathbb{1}_A \mathbb{1}_B$.

Furthermore if (A_n) is a sequence of disjoint sets in Σ and $A = \bigcup_{n=1}^{\infty} A_n$, show that $\mathbb{1}_A = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$.

Hint: In (a) consider what happens to both sides in each of the four cases: $x \in A, x \in B; x \in A, x \notin B$, etc.

ps:lils **2.2** Let (a_n) and (b_n) be bounded sequences of real numbers. Show that

(a) $\limsup_{n\to\infty} a_n = -\liminf_{n\to\infty}(-a_n)$,

(b) $\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$,

(c) $\liminf_{n\to\infty}(a_n + b_n) \ge \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$,

(d) if $a_n, b_n \ge 0$ for all $n \in \mathbb{N}$, $\limsup_{n\to\infty}(a_n b_n) \le (\limsup_{n\to\infty} a_n)(\limsup_{n\to\infty} b_n)$,

(e) if $a_n, b_n \ge 0$ for all $n \in \mathbb{N}$, $\liminf_{n\to\infty}(a_n b_n) \ge (\liminf_{n\to\infty} a_n)(\liminf_{n\to\infty} b_n)$,

(f) $\limsup_{n\to\infty} |a_n| = 0 \Rightarrow (a_n)$ converges to 0.

nstants_meas **2.3** Let (S, Σ) be a measurable space and $f : S \to \mathbb{R}$ be a constant function, i.e. there exists $c \in \mathbb{R}$ so that $f(x) = c$ for all $x \in S$. Show that f is measurable.

pen_int_meas **2.4** If (S, Σ) be a measurable space and $f : S \to \mathbb{R}$ show that f is measurable if and only if $f^{-1}((a, b)) \in \Sigma$ for all $-\infty \le a < b \le \infty$.

st_meas_func **2.5** Let (S, Σ) be a measurable space and $f : S \to \mathbb{R}$ be a measurable function.

(a) Show that $g = f + c$ is measurable, where $c \in \mathbb{R}$ is fixed,

(b) Show that $g = kf$ is measurable, where $k \in \mathbb{R}$ is fixed.

compose_meas **2.6** Let (S, Σ) be a measurable space and $f : S \to \mathbb{R}$ be a measurable function. If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, show that $g \circ f$ is measurable from S to $\mathbb{R}$. What does this result tell us about functions of random variables in probability theory?

anslate_meas **2.7** Let $f : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Show that the mapping $h : \mathbb{R} \to \mathbb{R}$ is measurable, where $h(x) = f(x + y)$ for all $x \in \mathbb{R}$, and where $y \in \mathbb{R}$ is fixed.

ps:abs_meas **2.8** Let (S, Σ) be a measurable space and $f : S \to \mathbb{R}$ be a function. Define the function $|f| : S \to \mathbb{R}$ by $|f|(x) = |f(x)|$ for all $x \in \mathbb{R}$. Show that

(a) $f = f_+ - f_-$,

(b) $|f| = f_+ + f_-$,

(c) if f is measurable then $|f|$ is also measurable.

ps:diff_meas **2.9** Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Explain why both f and its derivative f' are measurable functions.

ps:mono_meas **2.10** Let $f : \mathbb{R} \to \mathbb{R}$ be monotonic increasing. Show that it is measurable.

ps:lims_meas **2.11** Let (S, Σ) be a measure space and (f_n) be a sequence of measurable functions from S to $\mathbb{R}$. Let $f : S \to \mathbb{R}$ be measurable. We say that $f_n \to f$ *almost everywhere* or (*a.e.*) for short, if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in S - A$ where $m(A) = 0$. Show that if $f_n \to f$ (a.e.) and $g_n \to g$ (a.e.) then

(a) $f_n^2 \to f^2$ (a.e.)

(b) $f_n + g_n \to f + g$ (a.e.)

(c) $f_n g_n \to fg$ (a.e.)

ps:open_sets **2.12** Recall the definition of an open set, from Section 2.2.

(a) Let O_1 and O_2 be open subsets of $\mathbb{R}$. Show that $O_1 \cup O_2$ and $O_1 \cap O_2$ are also open.

(b) For each $n \in \mathbb{N}$ let O_n be an open subset of $\mathbb{R}$. Consider the following claims:

1. $A = \bigcup_{n\in\mathbb{N}} O_n$ is open.
2. $B = \bigcap_{n\in\mathbb{N}} O_n$ is open.

Which of these claims are true? Give a proof or a counterexample in each case.

(c) A set $C \subseteq \mathbb{R}$ is said to be *closed* if $\mathbb{R} \setminus C$ is open. Which of your results from parts (a) and (b) hold for closed sets?

Challenge questions

ps:upper_sc **2.13** A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *upper-semicontinuous* at $x \in \mathbb{R}$, if given any $\epsilon > 0$ there exists $\delta > 0$ so that $f(y) < f(x) + \epsilon$ whenever $|x - y| < \delta$.

(a) Show that $f = \mathbb{1}_{[a,\infty)}$ (where $a \in \mathbb{R}$) is upper-semicontinuous for all $x \in \mathbb{R}$,

(b) Deduce that the floor function $f(x) = \lfloor x \rfloor$, which is equal to the greatest integer less than or equal to x, is upper-semicontinuous at all $x \in \mathbb{R}$.

(c) Show that if f is upper-semicontinuous for all $x \in \mathbb{R}$ then f is measurable.

Chapter 3

Lebesgue Integration

_integration

sec:leb_intro

3.1 Introduction

The concept of integration as a technique that both acts as an inverse to the operation of differentiation and also computes areas under curves goes back to the origin of the calculus and the work of Isaac Newton (1643-1727) and Gottfried Leibniz (1646-1716). It was Leibniz who introduced the $\int \cdots dx$ notation. The first rigorous attempt to understand integration as a limiting operation within the spirit of analysis was due to Bernhard Riemann (1826-1866). The approach to *Riemann integration* that is often taught (as in MAS221) was developed by Jean-Gaston Darboux (1842-1917). At the time it was developed, this theory seemed to be all that was needed but as the 19th century drew to a close, some problems appeared:

- One of the main tasks of integration is to recover a function f from its derivative f'. But some functions were discovered for which f' existed and was bounded, but was not Riemann integrable.
- Suppose (f_n) is a sequence of functions converging pointwise to f. The Riemann integral could not be used to find conditions for which
$$\int f(x)dx = \lim_{n\to\infty} \int f_n(x)dx. \tag{3.1}$$
{eq:int_conv

  Problem **3.17** illustrates some of the difficulties here; it gives an example of f_n, f such that $f_n(x) \to f(x)$ for all x, but in which (3.1) fails.
- Riemann integration was limited to computing integrals over $\mathbb{R}^n$ with respect to Lebesgue measure. Although it was not yet apparent, the emerging theory of probability would require the calculation of expectations of random variables X: $\mathbb{E}(X) = \int_\Omega X(\omega)d\mathbb{P}(\omega)$.

A new approach to integration was needed. In this chapter, we'll study *Lebesgue integration*, which allow us to investigate $\int_S f(x)dm(x)$ where $f : S \to \mathbb{R}$ is a "suitable" measurable function defined on a general measure space (S, Σ, m).[1] If we take m to be Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we recover the familiar integral $\int_\mathbb{R} f(x)dx$ but we will now be able to integrate many more functions (at least in principle) than Riemann and Darboux. If we take X to be a random variable on a probability space, we get its expectation $\mathbb{E}(X)$.

[1] We may also integrate extended measurable functions, but will not develop that here.

Notation. For simplicity we usually write $\int_S f dm$ instead of $\int_S f(x) dm(x)$. To simplify even further we'll sometimes write $I(f) = \int_S f dm$. Note that many authors use $\int_S f(x) m(dx)$, with the same meaning. In French textbooks they often write $\int_S dm\, f$.

3.2 The Lebesgue Integral for Simple Functions

We'll present the construction of the Lebesgue integral in three steps: Step 1: Indicator functions, Step 2: Simple Functions, Step 3: Non-negative measurable functions, Step 4: Integrable functions. The first two steps begin here.

Step 1. **Indicator Functions**
This is very easy and yet it is very important:
If $f = \mathbb{1}_A$ where $A \in \Sigma$

$$\int_S \mathbb{1}_A dm = m(A). \qquad (3.2)$$ {Lebind}

Step 2. **Simple Functions**
Let $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ be a non-negative simple function so that $c_i \geq 0$ for all $1 \leq i \leq n$. We extend (3.2) by linearity, i.e. we define

$$\int_S f dm = \sum_{i=1}^n c_i m(A_i), \qquad (3.3)$$ {Lebsim}

and note that $\int_S f dm \in [0, \infty]$.

Remark 3.2.1 *We can represent f in more than one way as a simple function. For example if $f = \mathbb{1}_{[0,1]}$ then also $f = \mathbb{1}_{[0,\frac{1}{2})} + \mathbb{1}_{[\frac{1}{2},1]}$. It is easy to guess that the value of* (3.3) *doesn't depend on the choice of representation. We omit a formal proof of this fact.*

In each step, we'll establish some useful properties of the integral. Because we expand the amount of functions we can integrate at each step, this will mean that we carry several properties with us as we go, and do some work in each step to upgrade them. We begin this process with the next theorem.

thm:Ls1 **Theorem 3.2.2** *If f and g are non-negative simple functions and $\alpha, \beta \geq 0$ then*

1. *("Linearity") $\int_S (\alpha f + \beta g) dm = \alpha \int_S f dm + \beta \int_S g dm$,*
2. *(Monotonicity) If $f \leq g$ then $\int_S f dm \leq \int_S g dm$.*

PROOF:

1. Let $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}, g = \sum_{j=1}^m d_j \mathbb{1}_{B_j}$. Since $\bigcup_{i=1}^n A_i = \bigcup_{j=1}^m B_j = S$, we have

$$f = \sum_{i=1}^n c_i \mathbb{1}_{A_i \cap S} = \sum_{i=1}^n c_i \mathbb{1}_{A_i \cap \bigcup_{j=1}^m B_j} = \sum_{i=1}^n \sum_{j=1}^m c_i \mathbb{1}_{A_i \cap B_j}.$$

   Here, the last equality follows by Problem 2.1 part (d). It follows that

$$\alpha f + \beta g = \sum_{i=1}^n \sum_{j=1}^m (\alpha c_i + \beta d_j) \mathbb{1}_{A_i \cap B_j}.$$

Thus

$$\begin{aligned}
I(\alpha f+\beta g) &= \sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha c_i+\beta d_j)m(A_i\cap B_j)\\
&= \alpha\sum_{i=1}^{n}c_i\sum_{j=1}^{m}m(A_i\cap B_j)+\beta\sum_{j=1}^{n}d_i\sum_{i=1}^{m}m(A_i\cap B_j)\\
&= \alpha\sum_{i=1}^{n}c_i m\left(A_i\cap\bigcup_{j=1}^{m}B_j\right)+\beta\sum_{j=1}^{m}d_j m\left(\bigcup_{i=1}^{n}A_i\cap B_j\right)\\
&= \alpha\sum_{i=1}^{n}c_i m(A_i\cap S)+\beta\sum_{j=1}^{m}d_j m(B_j\cap S)\\
&= \alpha\sum_{i=1}^{n}c_i m(A_i)+\beta\sum_{j=1}^{m}d_j m(B_j)\\
&= \alpha I(f)+\beta I(g).
\end{aligned}$$

2. By (1), $I(g)=I(f)+I(g-f)$ but $g-f$ is a non-negative simple function and so $I(g-f)\geq 0$. The result follows.

■

Definition 3.2.3 If $A\in\Sigma$, whenever $\int_S f dm$ is defined for some $f:S\to\mathbb{R}$ we define

$$I_A(f)=\int_A f dm=\int_S \mathbb{1}_A f dm.$$

In general there is no guarantee that $I_A(f)$ is defined for some function f! We need f to be one of the types of functions that we work with in the steps used to define the integral.

3.3 The Lebesgue Integral for Non-negative Measurable Functions

sec:leb_2

We haven't done any analysis yet and at some stage we need to make integrals interact nicely with limits! If f is measurable and non-negative, it may seem attractive to try to take advantage of Theorem 2.4.1 and define "$\int_S f dm = \lim_{n\to\infty} \int_S s_n dm$". But there are many different choices of simple functions that we could take to make an approximating sequence, and this risks making the limiting integral depend on that choice, which is undesirable. Lebesgue's key idea was to use the weaker notion of the supremum to 'approximate f from below' as follows:

Step 3. **Non-negative measurable functions**

$$\int_S f dm = \sup\left\{\int_S s dm, s \text{ simple}, 0 \le s \le f\right\}. \tag{3.4}$$ {Lebgen1}

With this definition, $\int_S f dm \in [0,\infty]$. We allow the possibility that this integral may be equal to $+\infty$. The set over which we take the supremum is non-empty by Theorem 2.5.1.

The use of the sup makes it easy to prove some properties but hard to prove others. We'll have to postpone a full proof of linearity until the next section, but here are some other properties that can be proved fairly easily.

thm:bprops

Theorem 3.3.1 *If $f, g : S \to \mathbb{R}$ are non-negative measurable functions,*

1. *(Monotonicity) If $f \le g$ then $\int_S f dm \le \int_S g dm$.*
2. *$\int_S \alpha f\, dm = \alpha \int_S f\, dm$ for all $\alpha \ge 0$,*
3. *If $A, B \in \Sigma$ with $A \subseteq B$ then $I_A(f) \le I_B(f)$,*
4. *If $A \in \Sigma$ with $m(A) = 0$ then $I_A(f) = 0$.*

PROOF: For (1),

$$\begin{aligned}\int_S f dm &= \sup\left\{\int_S s dm\,;\, s \text{ is simple, } 0 \le s \le f\right\}\\ &\le \sup\left\{\int_S s dm\,;\, s \text{ is simple, } 0 \le s \le g\right\}\\ &= \int_S g dm.\end{aligned}$$

Parts (2), (3) and (4) are Problem 3.4. ■

:markov_ineq

Lemma 3.3.2 (Markov's inequality) *If $f : S \to \mathbb{R}$ is a non-negative measurable function and $c > 0$.*

$$m(\{x \in S; f(x) \ge c\}) \le \frac{1}{c}\int_S f dm$$

PROOF: Let $E = \{x \in S; f(x) \ge c\}$. Note that $E = f^{-1}([c,\infty)) \in \Sigma$ as f is measurable (from Theorem 2.2.1, in the context of the extended real numbers). By Theorem 3.3.1 parts (1) and (2),

$$\int_S f\, dm \ge \int_S \mathbb{1}_E f\, dm \ge \int_S c\mathbb{1}_E\, dm = c\int_S \mathbb{1}_E\, dm = cm(E),$$

and the result follows. ■

Definition 3.3.3 Let $f, g : S \to \mathbb{R}$ be measurable. We say that $f = g$ *almost everywhere*, and write this for short as $f = g$ a.e., if

$$m(\{x \in S; f(x) \neq g(x)\} = 0.$$

In Problem **3.10** you can show that this gives rise to an equivalence relation on the set of all measurable functions. You should think of '$g = f$ a.e.' as saying 'as far as measure theory is concerned, f and g are might as well be equal'.

cor:mezero **Corollary 3.3.4** *If f is a non-negative measurable function and $\int_S f\,dm = 0$ then $f = 0$ (a.e.)*

PROOF: Let $A = \{x \in S; f(x) \neq 0\}$ and for each $n \in \mathbb{N}, A_n = \{x \in S; f(x) \geq 1/n\}$. Since $A = \bigcup_{n=1}^{\infty} A_n$, we have $m(A) \leq \sum_{n=1}^{\infty} m(A_n)$ by Theorem 1.5.2, and its sufficient to show that $m(A_n) = 0$ for all $n \in \mathbb{N}$. But by Markov's inequality $m(A_n) \leq n \int_S f\,dm = 0$. ■

3.3.1 Integration as a measure

Integrals of non-negative measurable functions give us a way of constructing measures. More precisely, it gives us a way of taking an existing measure m and generating a large family of new measures from m. In turn, knowing that integrals (or non-negative simple functions) are measure will allow us to deduce some more properties about integrals.

thm:fmeas **Theorem 3.3.5** *If $f : S \to \mathbb{R}$ is a non-negative measurable function, then the map $I_A : \Sigma \to [0,\infty]$ given by*

$$I_A(f) = \int_A f\,dm \tag{3.5}$$ {eq:IAf}

is a measure.

Proof: We must check properties (M1) and (M2) of Definition 1.3.1. We have $\int_\emptyset f\,dm = 0$ by Theorem 3.3.1 part (4), which establishes (M1). It remains to check (M2) i.e. that $I_A(f) = \sum_{n=1}^\infty I_{A_n}(f)$ whenever we have a disjoint union $A = \bigcup_{n=1}^\infty A_n$.

First, assume that $f = \mathbb{1}_B$ for some $B \in \Sigma$. Then by (3.2)

$$\begin{aligned} I_A(f) = m(B \cap A) &= m\left(B \cap \bigcup_{n=1}^\infty A_n\right) \\ &= \sum_{n=1}^\infty m(B \cap A_n) = \sum_{n=1}^\infty I_{A_n}(f), \end{aligned}$$

so the result holds in this case. You can then use linearity to show that it is true for non-negative simple functions.

Now let f be measurable and non-negative. Then by definition of the supremum, for any $\epsilon > 0$ there exists a simple function s with $0 \le s \le f$ so that $I_A(f) \le I_A(s) + \epsilon$. The result holds for simple functions and so by monotonicity we have

$$I_A(s) = \sum_{n=1}^\infty I_{A_n}(s) \le \sum_{n=1}^\infty I_{A_n}(f).$$

Combining this with the earlier inequality we find that

$$I_A(f) \le \sum_{n=1}^\infty I_{A_n}(f) + \epsilon.$$

But ϵ was arbitrary and so we conclude that

$$I_A(f) \le \sum_{n=1}^\infty I_{A_n}(f).$$

The second half of the proof will aim to establish the opposite inequality. First let $A_1, A_2 \in \Sigma$ be disjoint. Given any $\epsilon > 0$ we can, as above, find simple functions s_1, s_2 with $0 \le s_j \le f$, so that $I_{A_j}(s_j) \ge I_{A_j}(f) - \epsilon/2$ for $j = 1, 2$. Let $s = s_1 \vee s_2 = \max\{s_1, s_2\}$. Then s is simple (check this), $0 \le s \le f$ and $s_1 \le s, s_2 \le s$. So by monotonicity, $I_{A_j}(s) \ge I_{A_j}(f) - \epsilon/2$ for $j = 1, 2$. Add these two inequalities to find that

$$I_{A_1}(s) + I_{A_2}(s) \ge I_{A_1}(f) + I_{A_2}(f) - \epsilon.$$

But the result is true for simple functions and so we have

$$I_{A_1 \cup A_2}(s) \geq I_{A_1}(f) + I_{A_2}(f) - \epsilon.$$

By the definition (3.4), $I_{A_1 \cup A_2}(f) \geq I_{A_1 \cup A_2}(s)$ and so we have that

$$I_{A_1 \cup A_2}(f) \geq I_{A_1}(f) + I_{A_2}(f) - \epsilon.$$

But ϵ was arbitrary and so we conclude that $I_{A_1 \cup A_2}(f) \geq I_{A_1}(f) + I_{A_2}(f)$, which is the required inequality for unions of two disjoint sets. By induction we have

$$I_{A_1 \cup A_2 \cup \cdots \cup A_n}(f) \geq \sum_{i=1}^{n} I_{A_i}(f),$$

for any $n \geq 2$. But as $A_1 \cup A_2 \cup \cdots \cup A_n \subseteq A$ we can use Theorem 3.3.1 (3) to find that

$$I_A(f) \geq \sum_{i=1}^{n} I_{A_i}(f).$$

Now take the limit as $n \to \infty$ to deduce that $I_A(f) \geq \sum_{i=1}^{\infty} I_{A_i}(f)$, as was required. ■

Example 3.3.6 The *Gaussian measure* on $\mathbb{R}$ is obtained in this way by taking $f = \phi$ where $\phi : \mathbb{R} \to \mathbb{R}$ is given by $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and taking m as Lebesgue measure. We note an explicit connection with probability theory:

$$I_A(\phi) = \int_A \frac{1}{2\pi} e^{-x^2/2}\, dx$$

which you should recognize as equal to $\mathbb{P}[Z \in A]$ where $Z \sim N(0,1)$. Thus $A \mapsto I_A(f)$ is the law of a standard normal random variable. (Recall that normal random variables are also known as Gaussian random variables.)

cor:premct **Corollary 3.3.7** *Let $f : S \to \mathbb{R}$ be a non-negative measurable function and (E_n) be a sequence of sets in Σ with $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$. Set $E = \bigcup_{n=1}^{\infty} E_n$, Then*

$$\int_E f dm = \lim_{n \to \infty} \int_{E_n} f dm.$$

PROOF: By Theorem 3.3.5 the map $A \mapsto I_A(f)$ in (3.5) is a measure, so we can apply Theorem 1.7.1 to this measure and the increasing sequence of sets (E_n). This gives the required result. ■

e_ints_equal **Corollary 3.3.8** *If f and g are non-negative measurable functions and $f = g$ almost everywhere then $\int_S f\, dm = \int_s g\, dm$.*

PROOF: Let $A_1 = \{x \in S;\, f(x) = g(x)\}$ and $A_2 = \{x \in S;\, f(x) \neq g(x)\}$. Noting that $A_2 = \{x \in S;\, f(x) > g(x)\} \cup \{x \in S;\, f(x) < g(x)\}$ we have from Lemma 2.4.2 that $A_2 \in \Sigma$. Since $A_1 = S \setminus A_2$ we also have $A_1 \in \Sigma$. Hence $A_1, A_2 \in \Sigma$ with $A_1 \cup A_2 = S$, $A_1 \cap A_2 = \emptyset$ and $m(A_2) = 0$. So by Theorem 3.3.1 part (4), $\int_{A_2} f dm = \int_{A_2} g dm = 0$. But $\int_{A_1} f dm = \int_{A_1} g dm$ as $f = g$ on A_1 and so by Theorem 3.3.5,

$$\begin{aligned}\int_S f dm &= \int_{A_1} f dm + \int_{A_2} f dm \\ &= \int_{A_1} g dm + \int_{A_2} g dm = \int_S g dm.\end{aligned}$$

■

3.4 The Monotone Convergence Theorem

We haven't yet proved that $\int_S (f+g)dm = \int_S f dm + \int_S g dm$. Nor have we extended the integral beyond non-negative measurable functions. Before we can do either of these, we need to establish the monotone convergence theorem. This is the first of two important results that show the superiority of Lebesgue integration over Riemann integration, in that Lebesgue integration interacts nicely with limits.

We say that a sequence (f_n) be of measurable functions is *monotone increasing* if $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$. Note that in this case the pointwise limit $f = \lim_{n\to\infty} f_n$ automatically exists, is non-negative and also measurable (by Theorem 2.4.4), and f may take values in $[0,\infty]$. Similarly, we say that (f_n) is *monotone decreasing* if $f_n \geq f_{n+1}$. We say that $f_n(x) \to f(x)$ *almost everywhere* if $\{x \in \mathbb{R}\,;\, f_n(x) \nrightarrow f(x)\}$ is a null set.

thm:mct **Theorem 3.4.1 (Monotone Convergence Theorem)** *Let f_n, f be functions from S to $\mathbb{R}$. Suppose that f_n is measurable and:*

1. *(f_n) is a monotone increasing, and each f_n is non-negative.*
2. *$f_n(x) \to f(x)$ almost everywhere.*

Then

$$\int_S f_n \, dm \to \int_S f \, dm$$

as $n \to \infty$.

PROOF: Since (f_n) is increasing, $\hat{f}(x) = \lim_{n\to\infty} f(x)$ exists for all $x \in \mathbb{R}$. We have $f(x) = \hat{f}(x)$ almost everywhere by our second assumption, which by Corollary 3.3.8 means that $\int_s f\,dm = \int_S \hat{f}\,dm$. Hence, in fact we may assume (without loss of generality, by using $\hat{f}$ in place of f) that $f_n(x) \to f(x)$ for *all* x. Thus f is measurable by Theorem 2.4.4.

As $f = \sup_{n\in\mathbb{N}} f_n$, by monotonicity (Theorem 3.3.1(1)), we have

$$\int_S f_1 dm \leq \int_S f_2 dm \leq \cdots \leq \int_S f dm.$$

Hence by monotonicity of the integrals, $\lim_{n\to\infty} \int_S f_n dm$ exists (as an extended real number) and

$$\lim_{n\to\infty} \int_S f_n dm \leq \int_S f dm.$$

We must now prove the reverse inequality. To simplify notation, let $a = \lim_{n\to\infty} \int_S f_n dm$. So we need to show that $a \geq \int_S f dm$. Let s be a simple function with $0 \leq s \leq f$ and choose $c \in \mathbb{R}$ with $0 < c < 1$. Our plan is to show that $a \geq c \int_S s\,dm$ and then take a sup over c and s.

For each $n \in \mathbb{N}$, let $E_n = \{x \in S; f_n(x) \geq cs(x)\}$, and note that $E_n \in \Sigma$ for all $n \in \mathbb{N}$ by Proposition 2.3.3. Since (f_n) is increasing, it follows that $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$.

We also have $\bigcup_{n=1}^{\infty} E_n = S$. To verify this last identity, note that if $x \in S$ with $s(x) = 0$ then $x \in E_n$ for all $n \in \mathbb{N}$ and if $x \in S$ with $s(x) \neq 0$ then $f(x) \geq s(x) > cs(x)$ and so for some n, $f_n(x) \geq cs(x)$, as $f_n(x) \to f(x)$ as $n \to \infty$, i.e. $x \in E_n$.

By Theorem 3.3.1(3) and (1), we have

$$a = \lim_{n\to\infty} \int_S f_n dm \geq \int_S f_n dm \geq \int_{E_n} f_n dm \geq \int_{E_n} csdm.$$

As this is true for all $n \in \mathbb{N}$, we find that

$$a \geq \lim_{n\to\infty} \int_{E_n} csdm.$$

But by Corollary 3.3.7 (since (E_n) is increasing), and Theorem 3.3.1(2),

$$\lim_{n\to\infty} \int_{E_n} csdm = \int_S csdm = c \int_S sdm,$$

and so we deduce that

$$a \geq c \int_S sdm.$$

But $0 < c < 1$ is arbitrary so taking e.g. $c = 1 - 1/k$ with $k = 2, 3, 4, \ldots$ and letting $k \to \infty$, we find that

$$a \geq \int_S sdm.$$

But the simple function s for which $0 \leq s \leq f$ was also arbitrary, so now take the supremum over all such s and apply (3.4) to get

$$a \geq \int_S fdm,$$

and the proof is complete. ■.

cor:simpappr **Corollary 3.4.2** *Let $f : S \to \mathbb{R}$ be measurable and non-negative. There exists an increasing sequence of simple functions (s_n) converging pointwise to f so that*

$$\lim_{n\to\infty} \int_S s_n dm = \int_S fdm. \tag{3.6}$$

{eq:simple_a

PROOF: Apply Theorem 3.4.1 to the sequence (s_n) constructed in Theorem 2.5.1. ■

We'll look at examples of calculating integrals, using tools like the monotone convergence theorem, in Section 3.8. For now we have a more pressing need to finish developing the Lebesgue integral, which we continue with in the next two sections.

Theorem 3.4.3 *Let $f, g : S \to \mathbb{R}$ be measurable and non-negative. Then*

$$\int_S (f+g)dm = \int_S fdm + \int_S gdm.$$

PROOF: By Theorem 2.5.1 we can find an increasing sequence of simple functions (s_n) that converges pointwise to f and an increasing sequence of simple functions (t_n) that converges pointwise to g. Hence $(s_n + t_n)$ is an increasing sequence of simple functions that converges pointwise to $f + g$. So by Theorem 3.4.1 and Theorem 3.2.2(1),

$$\begin{aligned}
\int_S (f+g)dm &= \lim_{n\to\infty} \int_S (s_n + t_n)dm \\
&= \lim_{n\to\infty} \int_S s_n dm + \lim_{n\to\infty} \int_S t_n dm \\
&= \int_S fdm + \int_S gdm.
\end{aligned}$$

■

3.5 Fatou's Lemma

A delicate convergence result can be obtained as a consequence of the monotone convergence theorem. We present this as a theorem, although it is always called *Fatou's lemma* after the French mathematician (and astronomer) Pierre Fatou (1878-1929). It tells us how $\liminf$ and $\int$ interact.

thm:Fatou **Theorem 3.5.1 (Fatou's Lemma)** *If (f_n) is a sequence of non-negative measurable functions from S to $\mathbb{R}$ then*

$$\liminf_{n\to\infty} \int_S f_n dm \geq \int_S \liminf_{n\to\infty} f_n dm$$

PROOF: Define $g_n = \inf_{k\geq n} f_k$. Then (g_n) is an increasing sequence which converges to $\liminf_{n\to\infty} f_n$. Now as $f_l \geq \inf_{k\geq n} f_k$ for all $l \geq n$, by monotonicity (Theorem 3.3.1(1)) we have that for all $l \geq n$

$$\int_S f_l dm \geq \int_S \inf_{k\geq n} f_k dm,$$

and so

$$\inf_{l\geq n} \int_S f_l dm \geq \int_S \inf_{k\geq n} f_k dm.$$

Now take limits on both sides of this last inequality and then apply the monotone convergence theorem (on the right hand side) to obtain

$$\begin{aligned}\liminf_{n\to\infty} \int_S f_n dm &\geq \lim_{n\to\infty} \int_S \inf_{k\geq n} f_k dm \\ &= \int_S \lim_{n\to\infty} \inf_{k\geq n} f_k dm \\ &= \int_S \liminf_{n\to\infty} f_n dm\end{aligned}$$

■

Note that we do not require (f_n) to be a bounded sequence, so $\liminf_{n\to\infty} f_n$ should be interpreted as an extended measurable function, as discussed at the end of Chapter 2. The corresponding result for $\limsup$, in which case the inequality is reversed, is known as the *reverse Fatou lemma* and can be found as Problem **3.13**. Fatou's lemma is useful for when handling sequences of functions (f_n) for which the pointwise limit does not exist, or is not known to exist.

3.6 Lebesgue Integrability

At last we are ready for the final step in the construction of the Lebesgue integral - the extension from non-negative measurable functions to a class of measurable functions that are real-valued.

Step 4. **Integrable functions.**

For the final step we first take f to be an arbitrary measurable function. We define the positive and negative parts of f, which we denote as f_+ and f_- respectively by:

$$f_+(x) = \max\{f(x), 0\}, \quad f_-(x) = \max\{-f(x), 0\}.$$

Both f_+ and f_- are measurable (by Corollary 2.3.2) and non-negative. We have

$$f = f_+ - f_-,$$

and using Step 3, we see that we can construct both $\int_S f_+ dm$ and $\int_S f_- dm$, noting that one or both might be equal to $+\infty$. If *at least one* of these is not equal to $+\infty$ then we define

$$\int_S f dm = \int_S f_+ dm - \int_S f_- dm,$$

which is then an extended real number $\in [-\infty, +\infty]$. It is important that we also have a way of restricting to a class of functions that avoids having infinite integrals:

Definition 3.6.1 We say that f is *integrable* if both $\int_S f_+ \, dm$ and $\int_S f_- \, dm$ exist and are finite.

Note that $|f| = f_+ + f_-$, where $|f|$ is defined pointwise as $|f(x)| = |f|(x)$. Therefore,

$$f \text{ is integrable} \quad \Leftrightarrow \quad |f| \text{ is integrable} \quad \Leftrightarrow \quad \int_S |f| \, dm < \infty.$$

We also have the useful inequality

$$\left| \int_S f dm \right| \leq \int_S |f| dm, \tag{3.7}$$ {eq:triangle

for all integrable f. This is a version of the familiar triangle inequality $|\sum a_n| \leq \sum |a_n|$ which applies to integrals. Proof of (3.7) is Problem **3.9** part (a).

We will show in Theorem 3.10.3 that, if $f : [a, b] \to \mathbb{R}$ is Riemann integrable, then f is also Lebesgue integrable, and in this case the value of the two integrals is equal. Consequently, you may use all the facts you already know about Riemann integration on $\mathbb{R}$ to evaluate integrals of the form $\int_a^b f(x) \, dx$. This includes integration by substitution, by parts, the Fundamental Theorem of Calculus, and so on.

We will begin to look at examples where we evaluate integrals in the next section. First, let us complete the process of establishing the key properties of the Lebesgue integral. The first two parts of the next theorem give linearity, and the final parts is monotonicity.

thm:basicsli **Theorem 3.6.2** *Suppose that f and g are integrable functions from S to $\mathbb{R}$.*

1. *If $c \in \mathbb{R}$ then cf is integrable and $\int_S cf dm = c \int_S f dm$,*
2. *$f + g$ is integrable and $\int_S (f + g) dm = \int_S f dm + \int_S g dm$,*
3. *If $f \leq g$ then $\int_S f dm \leq \int_S g dm$.*

PROOF: (1) and (3) are Problem **3.8**. For (2), we may assume that both f, g are not identically 0. The fact that $f+g$ is integrable if f and g are follows from Problem **3.9** part (b). To show that the integral of the sum is the sum of the integrals, we first need to consider six different cases (writing $h = f+g$) (i) $f \geq 0, g \geq 0, h \geq 0$, (ii) $f \leq 0, g \leq 0, h \leq 0$, (iii) $f \geq 0, g \leq 0, h \geq 0$, (iv) $f \leq 0, g \geq 0, h \geq 0$, (v) $f \geq 0, g \leq 0, h \leq 0$, (vi) $f \leq 0, g \geq 0, h \leq 0$. Case (i) is Theorem 3.4.2. We'll just prove (iii). The others are similar. If $h = f+g$ then $f = h + (-g)$ and this reduces the problem to case (i). Indeed we then have

$$\int_S f dm = \int_S (f+g) dm + \int_S (-g) dm,$$

and so by (1)

$$\int_S (f+g) dm = \int_S f dm - \int_S (-g) dm = \int_S f dm + \int_S g dm.$$

Now write $S = S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5 \cup S_6$, where S_i is the set of all $x \in S$ for which case (i) holds for $i = 1, 2, \ldots, 6$. These sets are disjoint and measurable and using linearity of Step 3 from Theorem 3.3.5 we have

$$\int_S (f+g) dm = \sum_{i=1}^{6} \int_{S_i} (f+g) dm = \sum_{i=1}^{6} \int_{S_i} f dm + \sum_{i=1}^{6} \int_{S_i} g dm = \int_S f dm + \int_S g dm,$$

as was required. ■

lem:dom_int

Lemma 3.6.3 *Let $g : S \to \mathbb{R}$ be integrable and suppose that $f : S \to \mathbb{R}$ is measurable with $|f(x)| \leq |g(x)|$ almost everywhere. Then f is integrable.*

PROOF: By part 1 of Theorem 3.3.1 we have $\int_S |f|\, dm \leq \int_S |g|\, dm$, which completes the proof. (Exercise: Why don't we use part 3 of Theorem 3.6.2 here?) ■

rem:leb_C

Remark 3.6.4 ($\star$) *The definition of the Lebesgue integral can be extended to complex valued functions. Let (S, Σ, m) be a measure space and $f : S \to \mathbb{C}$. We can always write $f = f_1 + i f_2$, where the real and imaginary parts are $f_i : S \to \mathbb{R}$ $(i = 1, 2)$. We say that f is measurable/integrable if both f_1 and f_2 are. When f is integrable, we may define*

$$\int_S f dm = \int_S f_1 dm + i \int_S f_2 dm.$$

*You can check that f is integrable if and only if $\int_S |f| dm < \infty$, where we now have $|f| = \sqrt{f_1^2 + f_2^2}$. See Problems **3.18**-**3.22** for applications to the Fourier transform.*

3.7 The Dominated Convergence Theorem

We now present the last of our convergence theorems, the famous *Lebesgue dominated convergence theorem* - an extremely powerful tool in both the theory and applications of modern analysis:

thm:DCT **Theorem 3.7.1 (Dominated Convergence Theorem)** *Let f_n, f be functions from S to $\mathbb{R}$. Suppose that f_n is measurable and:*

1. *There is an integrable function $g : S \to \mathbb{R}$ so that $|f_n| \le g$ for all $n \in \mathbb{N}$.*
2. *$f_n(x) \to f(x)$ almost everywhere.*

Then f is integrable and

$$\int_S f_n dm \to \int_S f dm$$

as $n \to \infty$.

Proof: For the same reason as in the proof of Theorem 3.4.1, we may assume that in fact $f_n(x) \to f(x)$ for all x. Thus f is measurable by Theorem 2.4.4. Note that we didn't assume explicitly that f_n is integrable - because this fact follows immediately from the first assumption and Lemma 3.6.3.

Since $f_n(x) \to f(x)$ almost everywhere, $|f_n(x)| \to |f(x)|$ almost everywhere. By Fatou's lemma (Theorem 3.5.1) and monotonicity (Theorem 3.6.2 part (3)), we have

$$\begin{aligned}\int_S |f| dm &= \int_S \liminf_{n\to\infty} |f_n| dm \\ &\le \liminf_{n\to\infty} \int_S |f_n| dm \\ &\le \int_S g dm < \infty,\end{aligned}$$

and so f is integrable.

Also for all $n \in \mathbb{N}, g + f_n \ge 0$ so by Fatou's lemma again,

$$\int_S \liminf_{n\to\infty} (g + f_n) dm \le \liminf_{n\to\infty} \int_S (g + f_n) dm.$$

But $\liminf_{n\to\infty}(g + f_n) = g + \lim_{n\to\infty} f_n = g + f$ and (using Theorem 3.6.2(2)) $\liminf_{n\to\infty} \int_S (g + f_n) dm = \int_S g dm + \liminf_{n\to\infty} \int_S f_n dm$. We then conclude that

$$\int_S f dm \le \liminf_{n\to\infty} \int_S f_n dm. \tag{3.8}$$ {eq:dct_1}

Repeat this argument with $g + f_n$ replaced by $g - f_n$ which is also non-negative for all $n \in \mathbb{N}$. We then find that

$$-\int_S f dm \le \liminf_{n\to\infty} \left(-\int_S f_n dm\right) = -\limsup_{n\to\infty} \int_S f_n dm,$$

and so

$$\int_S f dm \ge \limsup_{n\to\infty} \int_S f_n dm \tag{3.9}$$ {eq:dct_2}

Combining (3.8) and (3.9) we see that

$$\limsup_{n\to\infty} \int_S f_n dm \le \int_S f dm \le \liminf_{n\to\infty} \int_S f_n dm \tag{3.10}$$ {eq:dct_3}

but we always have $\liminf_{n\to\infty}\int_S f_n dm \leq \limsup_{n\to\infty}\int_S f_n dm$ and so $\liminf_{n\to\infty}\int_S f_n dm = \limsup_{n\to\infty}\int_S f_n dm$. Then by Theorem 2.1.1 $\lim_{n\to\infty}\int_S f_n dm$ exists, and from (3.10) we deduce that $\int_S f dm = \lim_{n\to\infty}\int_S f_n dm$. ■

Corollary 3.7.2 (Bounded Convergence Theorem) *Let (S, Σ, m) be a finite measure space. For all $n \in \mathbb{N}$, let $f_n : S \to \mathbb{R}$ be measurable and converging pointwise as $n \to \infty$ to $f : S \to \mathbb{R}$. Suppose that there exists $K < \infty$ such that $|f_n(x)| \leq K$ for all $n \in \mathbb{N}, x \in S$. Then f is integrable and $\int_S dm f_n \to \int_S f\, dm$.*

PROOF: Set $g(x) = K$ for all x, note that $\int_S K\, dm = Km(S) < \infty$ and apply Theorem 3.7.1. ■

3.8 Calculations with the Lebesgue Integral

sec:int_calcs

Now lets look at a few examples of using the major results of Chapter 3 to integrate particular functions. We'll begin with the testing for integrability and using the monotone convergence theorem, before moving on to the bounded and dominated convergence theorems.

From now on, we'll often adopt a piece of common notation and write $\int \ldots dx$ for integration with respect to Lebesgue measure, which formally we would write as $\int \ldots \lambda(dx)$. We'll allow ourselves to do this in cases where its clear from the context that we mean to integrate with respect to Lebesgue measure.

ex:poly_int

Example 3.8.1 We aim show that $f(x) = x^{-\alpha}$ is integrable on $[1, \infty)$ for $\alpha > 1$.

For each $n \in \mathbb{N}$ define $f_n(x) = x^{-\alpha}\mathbb{1}_{[1,n]}(x)$. Then $(f_n(x))$ increases to $f(x)$ as $n \to \infty$. We have

$$\int_1^\infty f_n(x)dx = \int_1^n x^{-\alpha}dx = \frac{1}{\alpha-1}(1 - n^{1-\alpha}).$$

By the monotone convergence theorem,

$$\int_1^\infty x^{-\alpha}dx = \frac{1}{\alpha-1}\lim_{n\to\infty}(1 - n^{1-\alpha}) = \frac{1}{\alpha-1}.$$

ex:gamma_int

Example 3.8.2 We aim to show that $f(x) = x^\alpha e^{-x}$ is integrable on $[0, \infty)$ for $\alpha > 0$.

The key idea here is that e^{-x} tends to zero very quickly as $x \to \infty$, and we can use this fast convergence to overpower the 'opposing' fact that $x^\alpha \to \infty$. Recall that for any $M \geq 0$ we have $\lim_{x\to\infty} x^M e^{-x} = 0$, so that given any $\epsilon > 0$ there exists $R > 0$ so that $x > R \Rightarrow x^M e^{-x} < \epsilon$, and choose M so that $M - \alpha > 1$. Now write

$$x^\alpha e^{-x} = x^\alpha e^{-x}\mathbb{1}_{[0,R]}(x) + x^\alpha e^{-x}\mathbb{1}_{(R,\infty)}(x).$$

By part 2 of Theorem 3.6.2 here, the sum of two integrable functions is an integrable function, so we'll aim to prove that both terms on the right hand side are integrable. The first term on the right clearly is, because it is bounded on $[0, R]$ and zero elsewhere. For the second term we use that fact that for all $x > R$,

$$x^\alpha e^{-x} = x^M e^{-x}.x^{\alpha-M} < \epsilon x^{\alpha-M},$$

and which is thus integrable by Example 3.8.1. So the result follows by Lemma 3.6.3.

Remark 3.8.3 *Let $I \subseteq \mathbb{R}$ be an interval and recall $(I, \mathcal{B}(I), \lambda_I)$ from Remark 1.5.3, which are the restriction of the Borel sets and Lebesgue measure to the interval I. A similar restriction applies for integrals, that is if $f : \mathbb{R} \to \mathbb{R}$ then $\int_\mathbb{R} \mathbb{1}_I(x)f(x)\,d\lambda(x) = \int_I f(x)\,d\lambda_I(x)$, or in more standard notation if $I = [a, b]$ (for example) then*

$$\int_a^b f(x)\,dx = \int_\mathbb{R} \mathbb{1}_{[a,b]}(x)f(x)\,dx.$$

We often treat this result as so obvious that we use it for free, and you may do so within this course. We'll do so in the next example.

Strictly, it does requires a proof, which can be done by first proving the result explicitly for simple functions, then applying the monotone convergence theorem to extend to non-negative measurable functions, and lastly transferring to integrable functions. As can be seen from the proof, there is nothing special about intervals here: in fact I can be be any Borel set.

Example 3.8.4 We want to calculate

$$\lim_{n\to\infty}\int_0^1 \frac{nx^2}{nx+5}\,dx.$$

We'll work in the measure space $([0,1],\mathcal{B}([0,1]),\lambda)$ and consider the sequence of functions (f_n) where $f_n(x)=\frac{nx^2}{nx+5}$ for all $x\in[0,1], n\in\mathbb{N}$. Each f_n is continuous, hence measurable by Corollary 2.3.2. It is straightforward to check that $\lim_{n\to\infty} f_n(x)=x$ for all $x\in[0,1]$ and that $|f_n(x)|\le 1$ for all $n\in\mathbb{N}, x\in[0,1]$. So in this case, we can take $K=1$, and apply the Bounded Convergence Theorem to deduce that $f(x)=x$ is integrable, and

$$\lim_{n\to\infty}\int_{[0,1]}\frac{nx^2}{nx+5}\,dx=\int_{[0,1]} x\,dx$$

and we finish by evaluating $\int_{[0,1]} x\,dx=\left[\frac{x^2}{2}\right]_0^1=\frac{1}{2}$.

ex:series

Example 3.8.5 Summation of series is a special case of Lebesgue integration. Suppose that we are interested in $\sum_{n=1}^\infty a_n$, where $a_n\ge 0$ for all $n\in\mathbb{N}$. We consider the sequence (a_n) as a function $a:\mathbb{N}\to[0,\infty)$. We work with the measure space $(\mathbb{N},\mathcal{P}(\mathbb{N}),m)$ where m is counting measure. Then every sequence (a_n) gives rise to a non–negative measurable function a and

$$\sum_{n=1}^\infty a_n=\int_{\mathbb{N}} a(n)\,dm(n),$$

which is for you to show in Problem 3.7. The same formula holds for general $(a_n)\subseteq\mathbb{R}$ provided that $\sum_n |a_n|<\infty$ i.e. that a is integrable.

The monotone and dominated converge theorems now provide tools for working with *sequences of series*. For example, suppose that for each $m\in\mathbb{N}$ we have a sequence $(a_m(n))_{n\in\mathbb{N}}$, given by

$$a_m(n)=\frac{1}{n^3}+\frac{1}{1+m^2n^2}.$$

We can't easily compute the value of $\sum_{n\in\mathbb{N}} a_m(n)$ for any given m. But we can note that $a_m(n)\to\frac{1}{n^3}$ as $m\to\infty$, for all n, and that $|a_m(n)|\le g(n)=\frac{1}{n^3}+\frac{1}{n^2}$. We know from analysis that $\sum_n\frac{1}{n^3}$ and $\sum_n\frac{1}{n^2}$ are both finite, so g is integrable – which in this setting is just the claim that $\sum_n g(n)<\infty$. So the Dominated Convergence Theorem applies, and we obtain

$$\lim_{m\to\infty}\sum_{n\in\mathbb{N}}\left(\frac{1}{n^3}+\frac{1}{1+m^2n^2}\right)=\sum_{n\in\mathbb{N}}\frac{1}{n^3}.$$

3.9 Fubini's Theorem and Function Spaces ($\star$)

This section is included for interest. It is marked with a ($\star$) and it is off-syllabus. However the first topic (Fubini's theorem) is covered in greater detail for MAS451/6352 within Chapter 6 *and that more extensive treatment is examinable for MAS451/3562.*

3.9.1 Fubini's Theorem ($\star$)

ec:fubini_350

Let (S_i, Σ_i, m_i) be two measure spaces[2]and consider the product space $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, m_1 \times m_2)$ as discussed in Section 1.8. We can consider integration of measurable functions $f : S_1 \times S_2 \to \mathbb{R}$ by the procedure that we've already discussed and there is nothing new to say about the definition and properties of $\int_{S_1 \times S_2} f d(m_1 \times m_2)$ when f is either measurable and non-negative (so the integral may be an extended real number) or when f is integrable (and the integral is a real number.)

From a practical point of view we would often like to calculate a double integral by writing it as a *repeated integral* so that we first integrate with respect to m_1 and then with respect to m_2 (or vice versa). *Fubini's theorem*, which we will state without proof, tells us that we can do this provided that f is integrable with respect to the product measure. It is named in honour of the Italian mathematician Guido Fubini (1879-1943).

thm:fubini

Theorem 3.9.1 (Fubini's Theorem) *Let f be integrable on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, m_1 \times m_2)$ so that $\int_{S_1 \times S_2} |f(x, y)|(m_1 \times m_2)(dx, dy) < \infty$. Then*

1. *The mapping $f(x, \cdot)$ is m_2-integrable, almost everywhere with respect to m_1,*
2. *The mapping $f(\cdot, y)$ is m_1-integrable, almost everywhere with respect to m_2,*
3. *The mapping $x \to \int_{S_2} f(x, y) m_2(dy)$ is equal almost everywhere to an integrable function on S_1,*
4. *The mapping $y \to \int_{S_1} f(x, y) m_1(dy)$ is equal almost everywhere to an integrable function on S_2,*
5. $$\begin{aligned}\int_{S_1 \times S_2} f(x, y)(m_1 \times m_2)(dx, dy) &= \int_{S_1} \left(\int_{S_2} f(x, y) m_2(dy) \right) m_1(dx) \\ &= \int_{S_2} \left(\int_{S_1} f(x, y) m_1(dx) \right) m_2(dy).\end{aligned}$$

3.9.2 Function Spaces ($\star$)

This section is aimed at those taking courses in functional analysis. An important application of Lebesgue integration is to the construction of Banach spaces $L^p(S, \Sigma, m)$ of equivalence classes of real-valued functions that agree a.e. and which satisfy the requirement

$$||f||_p = \left(\int_S |f|^p dm \right)^{\frac{1}{p}} < \infty,$$

where $1 \le p < \infty$. In fact $|| \cdot ||_p$ is a norm on $L^p(S, \Sigma, m)$, but only if $p \ge 1$. This is the reason why, in Section 5.2, we will only define L^p convergence for $p \ge 1$.

[2]Technically speaking, the measures should have an additional property called *σ–finiteness* for the main result below to be valid.

When $p = 2$ we obtain a Hilbert space with inner product:

$$\langle f, g \rangle = \int_S f g dm.$$

There is also a Banach space $L^\infty(S, \Sigma, m)$ where

$$||f||_\infty = \inf\{M \geq 0; |f(x)| \leq M \text{ a.e.}\}.$$

The same construction works with $\mathbb{C}$ in place of $\mathbb{R}$. These spaces play important roles in functional analysis and its applications, including partial differential equations, probability theory and quantum mechanics.

3.10 Riemann Integration ($\star$)

rie_leb_ints

In this section, our aim is to show that if a bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then it is measurable and Lebesgue integrable. Moreover, in this case the Riemann and Lebesgue integrals of f are equal. We state this result formally as Theorem 3.10.3.

Consequently, the Lebesgue integral on $\mathbb{R}$ is at least as powerful as the Riemann one – when we integrate over finite intervals. In fact, we can integrate *many* more functions using Lebesgue integration than we could using Riemann integration. For example, with Riemann integration we could not conclude that $\int_{[a,b]} \mathbb{1}_{\mathbb{R}\setminus\mathbb{Q}}(x)dx = (b-a)$, but with Lebesgue integration we can.

We begin by briefly revising the Riemann integral. Note that this whole section is marked with a ($\star$), meaning that it is off-syllabus. It will be discussed briefly in lectures.

3.10.1 The Riemann Integral ($\star$)

A partition $\mathcal{P}$ of $[a,b]$ is a set of points $\{x_0, x_1, \ldots, x_n\}$ with $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. Define $m_j = \inf_{x_{j-1} \le x \le x_j} f(x)$ and $M_j = \sup_{x_{j-1} \le x \le x_j} f(x)$. We underestimate by defining

$$L(f,\mathcal{P}) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}),$$

and overestimate by defining

$$U(f,\mathcal{P}) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}),$$

A partition $\mathcal{P}'$ is said to be a *refinement* of $\mathcal{P}$ if $\mathcal{P} \subset \mathcal{P}'$. We then have

$$L(f,\mathcal{P}) \le L(f,\mathcal{P}'), \quad U(f,\mathcal{P}') \le U(f,\mathcal{P}). \tag{3.11}$$ {ref}

A sequence of partitions $(\mathcal{P}_n)$ is said to be *increasing* if $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$ for all $n \in \mathbb{N}$.

Now define the *lower integral* $L_{a,b}f = \sup_{\mathcal{P}} L(f,\mathcal{P})$, and the *upper integral* $U_{a,b}f = \inf_{\mathcal{P}} U(f,\mathcal{P})$. We say that f is *Riemann integrable* over $[a,b]$ if $L_{a,b}f = U_{a,b}f$, and we then write the common value as $\int_a^b f(x)dx$. In particular, every continuous function on $[a,b]$ is Riemann integrable. The next result is very useful:

thm:Pfind **Theorem 3.10.1** *The bounded function f is Riemann integrable on $[a,b]$ if and only if for every $\epsilon > 0$ there exists a partition $\mathcal{P}$ for which*

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon. \tag{3.12}$$ {part}

If (3.12) holds for some $\mathcal{P}$, it also holds for all refinements of $\mathcal{P}$. A useful corollary is

cor:Pcor **Corollary 3.10.2** *If the bounded function f is Riemann integrable on $[a,b]$, then there exists an increasing sequence $(\mathcal{P}_n)$ of partitions of $[a,b]$ for which*

$$\lim_{n\to\infty} U(f,\mathcal{P}_n) = \lim_{n\to\infty} L(f,\mathcal{P}_n) = \int_a^b f(x)dx$$

PROOF: This follows from Theorem (3.10.1) by successively choosing $\epsilon = 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$. If the sequence $(\mathcal{P}_n)$ is not increasing, then just replace $\mathcal{P}_n$ with $\mathcal{P}_n \cup \mathcal{P}_{n-1}$ and observe that this can only improve the inequality (3.12). ∎

3.10.2 The Connection (⋆)

thm:RiL **Theorem 3.10.3** *If $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then it is Lebesgue integrable, and the two integrals coincide.*

PROOF: We use the notation λ for Lebsgue measure in this section. We also write $M = \sup_{x\in[a,b]} |f(x)|$ and $m = \inf_{x\in[a,b]} |f(x)|$.

Let $\mathcal{P}$ be a partition as above and define simple functions,

$$g_{\mathcal{P}} = \sum_{j=1}^{n} m_j \mathbb{1}_{(x_{j-1},x_j]}, \quad h_{\mathcal{P}} = \sum_{j=1}^{n} M_j \mathbb{1}_{(x_{j-1},x_j]}.$$

Consider the sequences (g_n) and (h_n) which correspond to the partitions of Corollary 3.10.2 and note that

$$L_n(f) = \int_{[a,b]} g_n d\lambda, \quad U_n f = \int_{[a,b]} h_n d\lambda,$$

where $U_n(f) = U(f, \mathcal{P}_n)$ and $L_n(f) = L(f, \mathcal{P}_n)$. Clearly we also have for each $n \in \mathbb{N}$,

$$g_n \le f \le h_n. \tag{3.13}$$

{order}

Since (g_n) is increasing (by (3.11)) and bounded above by M, it converges pointwise to a measurable function g. Similarly (h_n) is decreasing and bounded below by m, so it converges pointwise to a measurable function h. By (3.13) we have

$$g \le f \le h. \tag{3.14}$$

{order1}

Again since $\max_{n\in\mathbb{N}}\{|g_n|, |h_n|\} \le M$, we can use dominated convergence to deduce that g and h are both integrable on $[a,b]$ and by Corollary 3.10.2,

$$\int_{[a,b]} g d\lambda = \lim_{n\to\infty} L_n(f) = \int_a^b f(x)dx = \lim_{n\to\infty} U_n(f) = \int_{[a,b]} h d\lambda.$$

Hence we have

$$\int_{[a,b]} (h-g) d\lambda = 0,$$

and so by Corollary 3.3.1, $h(x) = g(x)$ (a.e.). Then by (3.14) $f = g$ (a.e.) and so f is measurable[3] and also integrable. So $\int_{[a,b]} f d\lambda = \int_{[a,b]} g d\lambda$, and hence we have

$$\int_{[a,b]} f d\lambda = \int_a^b f(x)dx.$$

■

[3] I'm glossing over a subtlety here. It is not true in general, that a function that is almost everywhere equal to a measurable function is measurable. It works in this case due to a special property of the Borel σ-field known as 'completeness'.

3.10.3 Discussion ($\star$)

An important caveat is that Theorem 3.10.3 only applies to *bounded* closed intervals. On unbounded intervals, there are examples of functions are Riemann integrable[4] but not Lebesgue integrable. One such example is $\int_0^\infty \frac{\sin x}{x}\,dx$. Crucially, $\frac{\sin x}{x}$ oscillates above and below 0 as $x\to\infty$, and the Riemann integral $\int_0^\infty \frac{\sin x}{x}\,dx = \lim_{X\to\infty}\int_0^X \frac{\sin x}{x}\,dx$ only exists because these oscillations cancel each other out. In Lebesgue integration this isn't allowed to happen, and $\frac{\sin x}{x}$ fails to be Lebesgue integrable because $\int_0^\infty |\frac{\sin x}{x}|\,dx = \infty$.

Let's discuss these ideas in the context of infinite series which, as we showed in Example 3.8.5, are a special case of the Lebesgue integral. That is,

$$\int_{\mathbb{N}} a_n\,d\#(n) = \sum_{n=1}^\infty a_n$$

where $a:\mathbb{N}\to\mathbb{R}$ is a sequence, and $\#$ is the counting measure on $\mathbb{N}$. Note that (a_n) is integrable if and only if $\sum_n |a_n| < \infty$, which is usually referred to as 'absolute convergence' in the context of infinite series. The key is that when infinite series are absolutely convergent they are much better behaved, as the following result shows. A 're-ordering' of a series simply means arranging its terms in a different order.

Theorem *Let (a_n) be a real sequence.*

1. *Suppose $\sum_{n=1}^\infty |a_n| = \infty$ and $a_n\to 0$. Then, for any $\alpha\in\mathbb{R}$, there is a re-ordering $b_n = a_{p(n)}$ such that $\sum_{i=1}^n b_n \to \alpha$.*
2. *Suppose $\sum_n |a_n| < \infty$. Then, for any re-ordering $b_n = a_{p(n)}$, we have $\sum_{n=1}^\infty a_n = \sum_n b_n \in \mathbb{R}$.*

Imagine if we allowed something similar to case 1 was allowed to happen in integration, and let us think about integration over $\mathbb{R}$. It would mean that re-ordering the x-axis values (e.g. swap $[0,1)$ with $[1,2)$ and so on) could change the value of $\int_{\mathbb{R}} f(x)\,dx$! This would be nonsensical, and mean that integration over $\mathbb{R}$ no longer had anything to do with 'area under the curve'. So we have to avoid it, and we do so by restricting to integrable functions. Only then can we find nice conditions for 'limit' theorems like the dominated convergence theorem.

[4] Strictly, we should say 'improperly' Riemann integrable.

3.11 Exercises

On integrals of step functions

ps:sf_int **3.1** Let $f:\mathbb{R}\to\mathbb{R}$ be defined as follows

$$f=\begin{cases}0 & \text{if } x<-2\\ 1, & \text{if } -2\le x<-1,\\ 0 & \text{if } -1\le x<0,\\ 2 & \text{if } 0\le x<1,\\ 1 & \text{if } 1\le x<2,\\ 0, & \text{if } x\ge 2\end{cases}$$

Write f explicitly as a simple function and calculate $\int_{\mathbb{R}} f(x)dx$.

f_int_finite **3.2** Let (S,Σ,m) be a measure space, $A\in\Sigma$ and f be a real-valued simple function defined on S. Show that $f\mathbb{1}_A$ is also a simple function, which is non-negative if f is. If f is non-negative, what constraint can you impose to ensure that $I_A(f)=I(f\mathbb{1}_A)$ is finite?

ps:f+f- **3.3** Let $f:\mathbb{R}\to\mathbb{R}$ be defined as follows

$$f=\begin{cases}0 & \text{if } x<-2\\ -1, & \text{if } -2\le x<-1,\\ 1 & \text{if } -1\le x<0,\\ -2 & \text{if } 0\le x<1,\\ 3 & \text{if } 1\le x<2,\\ 0, & \text{if } x\ge 2\end{cases}$$

Write down f_+ and f_- and confirm that they are non-negative simple functions. Calculate $\int_{\mathbb{R}} f_+(x)dx$ and $\int_{\mathbb{R}} f_-(x)dx$ and hence also $\int_{\mathbb{R}} f(x)dx$.

On integrals of non-negative functions

ps:T331 **3.4** Prove Theorem 3.3.1 parts (2) to (4).

ps:chebychev **3.5** Prove the following version of *Chebychev's inequality*. If $f:S\to\mathbb{R}$ is a measurable function and $c>0$ then

$$m(\{x\in S; |f(x)|\ge c\})\le\frac{1}{c^2}\int_S f^2 dm.$$

Formulate and prove a similar inequality where c^2 is replaced by c^p for $p\ge 1$.

Hint: Imitate the method of proof for Markov's inequality, stated as Lemma 3.3.2.

ositivity_Lp **3.6** Extend Corollary 3.3.4 as follows. Show that if f is a real valued measurable function for which $\int_S |f|^p dm=0$ for some $p\ge 1$ then $f=0$ almost everywhere.

3.7 Let $(a_n)_{n\in\mathbb{N}}$ be a real valued sequence, viewed as a function $a:\mathbb{N}\to\mathbb{R}$ with $a_n=a(n)$. We work over the measure space $(\mathbb{N},\mathcal{P}(\mathbb{N}),\#)$, where $\#$ denotes counting measure.

(a) Suppose that $a_n\ge 0$ and fix $N\in\mathbb{N}$. Let $a_n^{(N)}=\mathbb{1}_{\{n\le N\}}a_n$. Show that $a^{(N)}$ is a simple function, write down its integral, and use the monotone convergence theorem to deduce that

$$\int_{\mathbb{N}} a\,d\# = \sum_{n=1}^{\infty} a_n. \tag{3.15}$$ {eq:ps_serie

(b) Now consider a general $a = (a_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$. Explain briefly why a is integrable if and only if $\sum_n |a_n| < \infty$ and deduce that (3.15) holds in this case too.

s:series_int

On integrable functions and the convergence theorems

ps:T351 **3.8** Prove Theorem 3.6.2 parts (1) and (3).

Hint: For (1), consider the cases $c \geq 0, c = -1$ and $c < 0$ $(c \neq -1)$ separately.

ps:abs_int **3.9** Show that if f and g are integrable functions then

(a) $|\int_S f\,dm| \leq \int_S |f|\,dm$,

(b) $\int_S |f+g|\,dm \leq \int_S |f|\,dm + \int_S |g|\,dm$.

ps:ae_equiv **3.10** Show that $f = g$ (a.e.) defines an equivalence relation on the set of all real-valued measurable functions defined on (S, Σ, m).

nt_nonconv_1 **3.11** Consider the sequence (f_n) on the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ where $f_n = n\mathbb{1}_{(0,1/n)}$. Show that (f_n) converges pointwise to zero, but that $\int_{\mathbb{R}} f_n\,d\lambda = 1$ for all $n \in \mathbb{N}$.

Do any of the Monotone/Dominated Convergence Theorems, or Fatou's Lemma, apply to this situation?

_on_incr_set **3.12** Let (S, Σ, m) be a measure space and (A_n) be a sequence of disjoint sets with $A_n \in \Sigma$ for each $n \in \mathbb{N}$. define $A = \bigcup_{n=1}^{\infty} A_n$. Let $f : S \to \mathbb{R}$ be measurable. Show that $f\mathbb{1}_A$ is integrable if and only if $f\mathbb{1}_{A_n}$ is integrable for each $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} \int_{A_n} |f|dm < \infty$.

Hint: Use the monotone convergence theorem.

ps:rev_fatou **3.13** Prove the *reverse Fatou lemma*, i.e. if (f_n) is a sequence of non-negative measurable functions for which $f_n \leq f$ for all $n \in \mathbb{N}$ where f is integrable then

$$\limsup_{n\to\infty} \int_S f_n dm \leq \int_S \limsup_{n\to\infty} f_n dm.$$

Hint: Apply Fatou's lemma to $f - f_n$.

ps:cos_int **3.14** Show that if $f : \mathbb{R} \to \mathbb{R}$ is integrable then so are the mappings $x \to \cos(\alpha x)f(x)$ and $x \to \sin(\beta x)f(x)$, where $\alpha, \beta \in \mathbb{R}$. Deduce that

$$\lim_{n\to\infty} \int_{\mathbb{R}} \cos(x/n)f(x)dx = \int_{\mathbb{R}} f(x)dx.$$

ps:ints_cts **3.15** Let (S, Σ, m) be a measure space and $f : [a,b] \times S \to \mathbb{R}$ be a measurable function for which

(i) The mapping $x \to f(t,x)$ is integrable for all $t \in [a,b]$,

(ii) The mapping $t \to f(t,x)$ is continuous for all $x \in S$,

(iii) There exists a non-negative integrable function $g : S \to \mathbb{R}$ so that $|f(t,x)| \leq g(x)$ for all $t \in [a,b], x \in S$.

Use the dominated convergence theorem to show that the mapping $t \to \int_S f(t,x)dm(x)$ is continuous on $[a,b]$.

Hint: Use continuity in terms of sequences, that is show that $\lim_{n\to\infty} \int_S f(t_n,x)dm(x) = \int_S f(t,x)dm(x)$ for any sequence (t_n) satisfying $\lim_{n\to\infty} t_n = t$.

Challenge questions

ff_under_int **3.16** Let (S, Σ, m) be a measure space and $f : [a, b] \times S \to \mathbb{R}$ be a measurable function for which

(i) The mapping $x \to f(t, x)$ is integrable for all $t \in [a, b]$,

(ii) The mapping $t \to f(t, x)$ is differentiable for all $x \in S$,

(iii) There exists a non-negative integrable function $h : S \to \mathbb{R}$ so that $\left|\frac{\partial f(t,x)}{\partial t}\right| \le h(x)$ for all $t \in [a, b], x \in S$.

Show that the mapping $t \to \int_S f(t, x)dm(x)$ is differentiable on (a, b) and that

$$\frac{d}{dt}\int_S f(t,x)dm(x) = \int_S \frac{\partial f(t,x)}{\partial t}dm(x).$$

Hint: Use the mean value theorem.

:nonconv_int **3.17** Let

$$\begin{aligned} f(x) &= -2xe^{-x^2} \\ f_n(x) &= \sum_{r=1}^{n} \left(-2r^2xe^{-r^2x^2} + 2(r+1)^2xe^{-(r+1)^2x^2}\right) \end{aligned}$$

for all $x \in \mathbb{R}$.

(a) Show that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in \mathbb{R}$.

(b) Let $a > 0$. Show that f and f_n are Riemann integrable over $[0, a]$ for all $n \in \mathbb{N}$ but that

$$\int_0^a f(x)dx \neq \lim_{n\to\infty}\int_0^a f_n(x)dx.$$

Neither the monotone or dominated convergence theorems can be used here (follow up exercise: explain why not). This example illustrates that things can go badly wrong without them, even when $f_n(x) \to f(x)$ for all x.

Additional questions ($\star$)

These questions explore the definition and properties of the Fourier transform. They are off syllabus, but you may find them interesting. It may help to recall Remark 3.6.4.

urier_linear **3.18** If f is an integrable function on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where λ is Lebesgue measure, define its *Fourier transform* $\widehat{f}(y)$ for each $y \in \mathbb{R}$, by

$$\begin{aligned}\widehat{f}(y) &= \int_{\mathbb{R}} e^{-ixy} f(x)dx \\ &= \int_{\mathbb{R}} \cos(xy)f(x)dx - i\int_{\mathbb{R}} \sin(xy)f(x)dx.\end{aligned}$$

Prove that $|\widehat{f}(y)| < \infty$ and so $\widehat{f}$ is a well-defined function from $\mathbb{R}$ to $\mathbb{C}$. Show also that the Fourier transformation $\mathcal{F}f = \widehat{f}$ is linear, i.e. for all integrable f, g, and $a, b \in \mathbb{R}$ we have

$$\widehat{af + bg} = a\widehat{f} = b\widehat{g}.$$

fourier_jump **3.19** Recall Dirichlet's jump function $\mathbb{1}_{\mathbb{Q}}$. Does it make sense to write down the Fourier coefficients $a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbb{1}_{\mathbb{Q}}(x)\cos(nx)dx$ and $b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbb{1}_{\mathbb{Q}}(x)\sin(nx)dx$ as Lebesgue integrals? If so, what values do they have? Can you associate a Fourier series to $\mathbb{1}_{\mathbb{Q}}$? If so, (and if it is convergent) what does it converge to?

er_translate **3.20** Fix $a \in \mathbb{R}$ and define the shifted function $f_a(x) = f(x-a)$. If f is integrable, show that f_a also is, and deduce that $\widehat{f_a}(y) = e^{-iay}\widehat{f}(y)$ for all $y \in \mathbb{R}$.

:fourier_cts **3.21** Show that the mapping $y \to \widehat{f}(y)$ is continuous from $\mathbb{R}$ to $\mathbb{C}$.

Hint: In this question, and the next one, you may use the fact that Lebesgue's dominated convergence theorem continues to hold for complex valued functions, where $|\cdot|$ is interpreted as the usual modulus of complex numbers.

:fourier_xfx **3.22** Suppose that the mappings $x \to f(x)$ and $x \to xf(x)$ are both integrable. Show that $y \to \widehat{f}(y)$ is differentiable and that for all $y \in \mathbb{R}$,

$$(\widehat{f})'(y) = -i\widehat{g}(y),$$

where $g(x) = xf(x)$ for all $x \in \mathbb{R}$.

Hint: Use the inequality $|e^{ib} - 1| \leq |b|$ for $b \in \mathbb{R}$.

Remark 3.11.1 *Analogues of the results of Problems 3.18-3.22, with slight modifications, also hold for the Laplace transform $\mathcal{L}f(y) = \int_{[0,\infty)} e^{-yx}f(x)dx$, where $y \geq 0$ and $x \mapsto e^{-yx}f(x)$ is assumed to be integrable on $[0,\infty)$.*

Chapter 4

Probability and Measure

ob_with_meas

In this chapter we will examine probability theory from the measure theoretic perspective. The realisation that measure theory is the foundation of probability is due to the Russian mathematician A. N. Kolmogorov (1903-1987) who in 1933 published the hugely influential "Grundbegriffe der Wahrscheinlichkeitsrechnung" (in English: *Foundations of the Theory of Probability*). Since that time, measure theory has underpinned all mathematically rigorous work in probability theory and has been a vital tool in enabling the theory to develop both conceptually and in applications.

We have already noted that a probability is a measure, random variables are measurable functions and expectation is a Lebesgue integral – but it is not fair to claim that "probability theory" can be reduced to a subset of "measure theory". This is because in probability we model chance and unpredictability, which brings in a set of intuitions and ideas that go well beyond those of weights and measures. The Polish mathematician Mark Kac (1914-1984) famously described probability theory as "measure theory with a soul."

Definition 4.0.1 A measure m is said to be a probability measure if it has total mass 1.

4.1 Probability as Measure

Let us review what we know so far. In this chapter we will work with general probability spaces of the form $(\Omega, \mathcal{F}, \mathbb{P})$ where the *probability measure* $\mathbb{P}$ is a finite measure on $(\Omega, \mathcal{F})$ having total mass 1. So

$$\mathbb{P}[\Omega] = 1 \quad \text{and} \quad 0 \le \mathbb{P}[A] \le 1 \text{ for all } A \in \mathcal{F}.$$

Intuitively, $\mathbb{P}[A]$ is the probability that the event $A \in \mathcal{F}$ takes place. We will generally assign a special status to probability measures and expectations by writing their arguments in square brackets e.g. $\mathbb{P}[A]$ instead of $\mathbb{P}(A)$. This just a convention – there is no difference in mathematical meaning.

Since $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$, by (M2) we have $1 = \mathbb{P}[\Omega] = \mathbb{P}[A \cup A^c] = \mathbb{P}[A] + \mathbb{P}[A^c]$ so that

$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A].$$

A random variable X is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. If $A \in \mathcal{B}(\mathbb{R})$, it is standard to use the notation $(X \in A)$ to denote the event $X^{-1}(A) \in \mathcal{F}$. The *law* or *distribution* of X is the induced probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $p_X(B) = \mathbb{P}[X^{-1}(B)]$ for $B \in \mathcal{B}(\mathbb{R})$.

So

$$p_X(B) = \mathbb{P}[X \in B] = \mathbb{P}[X^{-1}(B)] = \mathbb{P}[\{\omega \in \Omega; X(\omega) \in B\}].$$

The *expectation* of X is the Lebesgue integral

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

which makes sense and yields a finite quantity if and only if X is *(Lebesgue) integrable* in the sense of Chapter 3. In this case, we write $\mu_X = \mathbb{E}(X)$ and call it the *mean* of X. Note that for all $A \in \mathcal{F}$

$$\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A]$$

because 1_A is a simple function $\mathbb{1}_A(\omega) = \mathbb{1}\{\omega \in A\}$.

By the result of Problem **2.6**, any Borel measurable function f from $\mathbb{R}$ to $\mathbb{R}$ enables us to construct a new random variable $f(X)$ for which $f(X)(\omega) = f(X(\omega))$ for all $\omega \in \Omega$. For example we may take $f(x) = x^n$ for all $n \in \mathbb{N}$. Then the nth moment $\mathbb{E}[X^n]$ will exist and be finite if $|X|^n$ is integrable. If X has a finite second moment then its *variance* $\mathrm{var}(X) = \mathbb{E}[(X - \mu)^2]$ always exists (see Problem **4.9**). It is common to use the notation $\sigma_X^2 = \mathrm{var}(X)$. The *standard deviation* of X is $\sigma_X = \sqrt{\mathrm{var}(X)}$. When it is clear which random variable we mean, we might write simply μ and σ in place of μ_X, σ_X.

Here's some useful notation. If X and Y are random variables defined on the same probability space and $A_1, A_2 \in \mathcal{B}(\mathbb{R})$ it is standard to write:

$$\mathbb{P}[X \in A_1, Y \in A_2] = \mathbb{P}[\{X \in A_1\} \cap \{Y \in A_2\}].$$

Let us now update the results of Section 1.7 into the language of probability. Recall that a sequence of sets (A_n) with $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ is *increasing* if $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$. Similarly, we say that a sequence (B_n) with $B_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ is *decreasing* if $B_n \supseteq B_{n+1}$ for all $n \in \mathbb{N}$.

one_events_P

Theorem 4.1.1 *We have:*

1. *Suppose (A_n) is increasing and $A = \bigcup_n A_n$. Then $\mathbb{P}[A] = \lim_{n\to\infty} \mathbb{P}[A_n]$.*
2. *Suppose (B_n) is decreasing and $B = \bigcap_n B_n$. Then $\mathbb{P}[B] = \lim_{n\to\infty} \mathbb{P}[B_n]$.*

PROOF: This is just Theorem 1.7.1 applied to probability measures. Note in particular that the condition of part 2 holds automatically here, because in probability all events (i.e. measurable sets) have finite measure. ■

The intuition for the above theorem should be clear. The set A_n gets bigger as $n \to \infty$ and, in doing so, gets ever closer to A; the same is true of their probabilities. Similarly for B_n, which gets smaller and closer to B. This result is a probabilistic analogue of the well known fact that monotone increasing (resp. decreasing) sequences of real numbers converge to the respective sups and infs.

4.2 The Cumulative Distribution Function

Let $X : \Omega \to \mathbb{R}$ be a random variable. Its *cumulative distribution function* or *cdf* is the mapping $F_X : \mathbb{R} \to [0,1]$ defined for each $x \in \mathbb{R}$ by

$$F_X(x) = \mathbb{P}[X \leq x] = p_X((-\infty, x]).$$

When X is clear from the context we might write F instead of F_X.

The next result gathers together some useful properties of the cdf. Recall that if $f : \mathbb{R} \to \mathbb{R}$, the *left limit* at x is $\lim_{y\uparrow x} f(y) = \lim_{y\to x, y<x} f(y)$, and the *right limit* at x is $\lim_{y\downarrow x} f(y) = \lim_{y\to x, y>x} f(y)$. In general, left and right limits may not exist, but they do if the function f is monotonic increasing (or decreasing).

thm:cdf **Theorem 4.2.1** *Let X be a random variable having cdf F.*

1. $\mathbb{P}[X > x] = 1 - F(x)$,
2. $\mathbb{P}[x < X \leq y] = F(y) - F(x)$ *for all* $x < y$.
3. *F is monotone increasing, i.e* $F(x) \leq F(y)$ *for all* $x < y$,
4. $\mathbb{P}[X = x] = F(x) - \lim_{y\uparrow x} F(y)$,
5. *The mapping $x \to F(x)$ is right continuous, i.e.* $F(x) = \lim_{y\downarrow x} F(y)$, *for all* $x \in \mathbb{R}$,
6. $\lim\limits_{x\to-\infty} F(x) = 0$ *and* $\lim\limits_{x\to\infty} F(x) = 1$.

Proof: Parts 1-3 are left as (easy!) exercises for you. Part 6 is Problem **4.1**.

Part 4: Let (a_n) be a sequence of positive numbers that decreases to zero. Let $x \in \mathbb{R}$ be arbitrary and for each $n \in \mathbb{N}$, define $B_n = (x - a_n < X \leq x)$. Then (B_n) decreases to the event $(X = x)$, so using part 2 and Theorem 4.1.1,

$$\mathbb{P}[X = x] = \lim_{n\to\infty} \mathbb{P}[B_n] = F(x) - \lim_{n\to\infty} F(x - a_n),$$

and the result follows.

Part 5: Let x and (a_n) be as in (4) and for each $n \in \mathbb{N}$ define $A_n = (X > x + a_n)$. The sets (A_n) are increasing to $(X > x)$, so using part 1 and Theorem 4.1.1 we find that

$$1 - F(x) = \lim_{n\to\infty} \mathbb{P}[A_n] = 1 - \lim_{n\to\infty} F(x + a_n),$$

and the result follows. ■

rem:cdf_cont **Remark 4.2.2** *By combining parts 4 and 5 of Theorem 4.2.1, we have that $\mathbb{P}[X = x_0] = 0$ if and only if $F_X(x)$ is continuous at $x = x_0$.*

Remark 4.2.3 *($\star$) It can be shown that a function $F : \mathbb{R} \to \mathbb{R}$ is the cdf of some random variable X if and only if it satisfies properties 3, 5 and 6 of Theorem 4.2.1.*

4.3 Discrete and Continuous Random Variables

You will probably recall that many useful random variables are found in two special cases. Formally, we say that a random variable X is a:

1. *continuous random variable* if its cdf F_X is continuous at every point $x \in \mathbb{R}$;
2. *discrete random variable* if F_X has jump discontinuities at a countable set of points and is constant between these jumps.

Note that if F_X is continuous at x then $\mathbb{P}(X = x) = 0$ Remark 4.2.2. In particular, this applies to all $x \in \mathbb{R}$ for continuous random variables. Many random random variables are neither discrete nor continuous.

We now point out a technicality that is often forgotten in less rigorous courses: a continuous random variable does not need to have a probability density function! Strictly speaking, those that do have a special name; we say X is a

3. *absolutely continuous random variable* if there exists an integrable function $f_X : \mathbb{R} \to \mathbb{R}$ so that $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$ for all $x \in \mathbb{R}$.

The function f_X is called the *probability density function* or *pdf* of X. Clearly $f_X \geq 0$ (a.e.) and by Theorem 4.2.1 (6) we have $\int_{-\infty}^{\infty} f_X(y)dy = 1$.

Lemma 4.3.1 *Every absolutely continuous random variable is a continuous random variable.*

PROOF: Note that

$$\int_{-\infty}^{x} f_X(y)\,dy = \int_{\mathbb{R}} \mathbb{1}_{(y\leq x)} f_X(y)\,dy.$$

We want to prove this is a continuous function of x, for which we'll use the dominated convergence theorem and the definition of continuity in terms of sequences. Let $(x_n) \subseteq \mathbb{R}$ be any sequence such that $x_n \to x$. Note that $|\mathbb{1}_{(y\leq x_n)} f_X(y)| \leq |f_X(y)|$ with $\int_{\mathbb{R}} |f_X(y)|\,dy = \mathbb{P}[X \in \mathbb{R}] = 1 < \infty$. If $y \neq x$ then we have $\mathbb{1}_{(y\leq x_n)} \to \mathbb{1}_{(y\leq x)}$ which means that $\mathbb{1}_{(y\leq x_n)} f_X(y) \to \mathbb{1}_{(y\leq x)} f_X(y)$ for almost all $y \in \mathbb{R}$. Hence by the dominated convergence theorem we have

$$\int_{-\infty}^{x_n} f_X(y)\,dy \to \int_{-\infty}^{x} f_X(y)\,dy$$

as $n \to \infty$, as required. ■

We have already seen the example of the Gaussian random variable that is absolutely continuous. Most useful continuous random variables are absolutely continuous; other examples that you may have encountered previously include the uniform, exponential, Student t, gamma and beta distributions. Typical examples of discrete random variables are the binomial, geometric and Poisson distributions.

4.4 Independence

sec:indep

In this subsection we consider the meaning of independence for *infinite sequences* of events and random variables. A useful heuristic is 'independence means multiply'. Recall that two events $A_1, A_2 \in \mathcal{F}$ are *independent* if

$$\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2].$$

For three events we would use $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3]$ and so on.

For many applications, we want to discuss independence of infinitely many events, or to be precise a sequence (A_n) of events with $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$. The definition of independence is extended from the finite case by considering all finite subsets of the sequence. Formally:

Definition 4.4.1 We say that the events in the sequence (A_n) are *independent* if the finite set $\{A_{i_1}, A_{i_2}, \ldots, A_{i_m}\}$ is independent for all finite subsets $\{i_1, i_2, \ldots, i_m\}$ of the natural numbers, i.e.

$$\mathbb{P}[A_{i_1} \cap A_{i_2} \cap \cdots, A_{i_m}] = \mathbb{P}[A_{i_1}]\mathbb{P}[A_{i_2}] \cdots \mathbb{P}[A_{i_m}].$$

Two random variables X and Y are said to be independent if $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]$ for all $A, B \in \mathcal{B}(\mathbb{R})$. This idea is extended to three or more random variables in the same way as above. For an infinite sequence of random variables (X_n), we say that the X_n are independent if every finite subset $X_{i_1}, X_{i_2}, \ldots, X_{i_m}$ of random variables is independent, i.e.

$$\mathbb{P}[X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \ldots, X_{i_m} \in A_{i_m}] = \mathbb{P}[X_{i_1} \in A_{i_1}]\mathbb{P}[X_{i_2} \in A_{i_2}] \cdots \mathbb{P}[X_{i_m} \in A_{i_m}]$$

for all $A_{i_1}, A_{i_2}, \ldots, A_{i_m} \in \mathcal{B}(\mathbb{R})$ and for all finite $\{i_1, i_2, \ldots, i_m\} \subseteq \mathbb{N}$.

We often want to consider random variables in $\mathbb{R}^d$, where $d \in \mathbb{N}$. Let us consider the case $d = 2$. A random variable in $\mathbb{R}^2$ $Z = (X, Y)$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ where $\mathcal{B}(\mathbb{R}^2)$ is the product σ-field introduced in Section 1.8. The law of Z is $p_Z = \mathbb{P} \circ Z^{-1}$, so that $p_Z(A) = \mathbb{P}[Z \in A]$ where $A \in \mathcal{B}(\mathbb{R}^2)$. The *joint law* of X and Y is $p_Z(A \times B) = \mathbb{P}[X \in A, Y \in B]$ for $A, B \in \mathcal{B}(\mathbb{R})$, and the *marginal laws* of X and Y are $p_X(A) = \mathbb{P}[X \in A]$ and $p_Y(B) = \mathbb{P}[Y \in B]$. From the definitions above, we have that X and Y are independent if and only if

$$p_Z(A \times B) = p_X(A)p_Y(B),$$

i.e. if the joint law factorises as the product of the two marginals. The same ideas extend to $\mathbb{R}^3$ with e.g. $W = (X, Y, Z)$ and so on.

thm:indep_E

Theorem 4.4.2 *Let X and Y be random variables.*

1. *If X and Y are independent and $f, g : \mathbb{R} \to \mathbb{R}$ are measurable functions then $f(X)$ and $g(Y)$ are independent.*
2. *If X and Y are independent and integrable then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*
3. *The following two conditions are equivalent:*
   (a) X and Y are independent;
   (b) $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ for all bounded measurable functions $f, g : \mathbb{R} \to \mathbb{R}$.

PROOF: The first part is left for you in Exercise **4.6**. For the second part,

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy p_Z(dx, dy) = \left(\int_{\mathbb{R}} x p_X(dx)\right)\left(\int_{\mathbb{R}} y p_Y(dy)\right) = \mathbb{E}[X]\mathbb{E}[Y]$$

Here, the first equality is the two-dimensional version of Problem **4.11**, and we have used Fubini's theorem (from Section 3.9.1) in the second equality to write the integral over $\mathbb{R}^2$ as a repeated integral. For the final part, recall that bounded random variables are integrable, so combining parts 1 and 2 gives that (a)⇒(b). To see that (b)⇒(a), take measurable sets $A, B \in \mathcal{B}(\mathbb{R})$ and set $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$. Then we have $\mathbb{E}[f(X)g(Y)] = \mathbb{P}[X \in A, Y \in B]$ and $\mathbb{E}[f(X)] = \mathbb{P}[X \in A]$, $\mathbb{E}[g(Y)] = \mathbb{P}[Y \in B]$, so (b) gives $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]$. ■

Remark 4.4.3 *Regarding part 1 of Theorem 4.4.2, note that* dependent *random variables X and Y can also satisfy $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. See Exercise 4.4 for an example of this.*

4.5 Exercises

:cdf_thm_bit **4.1** Use Theorem 4.1.1 to prove part 6 of Theorem 4.2.1.

rob_conv_thms **4.2** Write down 'probabilistic' versions of the monotone convergence theorem, Fatou's lemma and the dominated convergence theorem. That is, use random variables in place of measurable functions and expectation in place of the integral.

ps:EmaxXa **4.3** Let X be a random variable and $a \in \mathbb{R}$. Prove that

$$\mathbb{E}[\max\{X, a\}] \geq \max\{\mathbb{E}[X], a\}.$$

Hint: Write $\mathbb{E}[\max\{X, a\}]$ *as an integral.*

:indep_uncor **4.4** Let U be a random variable such that $\mathbb{P}[U = -1] = \mathbb{P}[U = 1] = \frac{1}{2}$ and let V be a random variable such that $\mathbb{P}[V = 0] = \mathbb{P}[V = 1] = \frac{1}{2}$, independent of U. Let $X = UV$ and $Y = U(1 - V)$. Show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ but that X and Y are *not* independent.

ps:indep_inf **4.5** (a) Let (A_n) be a sequence of independent events. Show that

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}} A_n\right] = \prod_{n=1}^{\infty} \mathbb{P}[A_n]. \tag{4.1}$$

{eq:inf_inde

(b) Recall that we define independence of a sequence of events (A_n) in terms of *finite* subsequences (e.g. as in Section 4.4). An 'obvious' alternative definition might to be use (4.1) instead. Why is this not a sensible idea?

s:indep_exts **4.6** (a) Let A and B be independent events. Show that their complements A^c and B^c are also independent.

(b) Let X and Y be independent random variables and $f, g : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Deduce that $f(X)$ and $g(Y)$ are also independent.

P_tail_bound **4.7** (a) Let X be a random variable that takes values in $\mathbb{N} \cup \{0\}$. Explain why $X = \sum_{i=1}^{\infty} \mathbb{1}_{\{X \geq i\}}$ and hence show that

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i].$$

(b) Let Y be a non-negative random variable. Deduce that

$$\sum_{k=1}^{\infty} \mathbb{P}[Y \geq k] \;\leq\; \mathbb{E}[Y] \;\leq\; 1 + \sum_{k=1}^{\infty} \mathbb{P}[Y \geq k].$$

:chebychev_2 **4.8** Let X be a real-valued random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Use the result of Exercise **3.5** to derive the probabilistic version of Chebychev's inequality:

$$\mathbb{P}\left[|X - \mu| \geq c\right] \leq \frac{\mathrm{var}(X)}{c^2},$$

where $\mathrm{var}(X)$ and $\mathbb{E}[X] = \mu$ are both assumed to be finite.

uchy_schwarz **4.9** (a) Suppose that X and Y are random variables and both X^2 and Y^2 are integrable. Prove the *Cauchy-Schwarz inequality*:

$$|\mathbb{E}[XY]| \leq (\mathbb{E}[X^2]^{\frac{1}{2}})(\mathbb{E}[Y^2]^{\frac{1}{2}}).$$

Hint: Consider $g(t) = \mathbb{E}[(X + tY)^2]$ *as a quadratic function of* $t \in \mathbb{R}$.

(b) Deduce that if X^2 is integrable then so is X, and in fact $|\mathbb{E}[X]|^2 \leq \mathbb{E}[X^2]$.

(c) Let X be *any* random variable with a finite mean $\mathbb{E}[X] = \mu$. Show that $\mathbb{E}[X^2] < \infty$ if and only if $\text{var}(X) < \infty$.

:exp_moments **4.10** (a) Let X be a non-negative random variable and $a > 0$. Show that $\mathbb{E}(e^{-aX}) \leq 1$.

(b) A random variable is said to have an *exponential moment* if $\mathbb{E}(e^{a|X|}) < \infty$ for some $a > 0$. Show that $X \sim N(0,1)$ has exponential moments for all $a > 0$.

(c) If X has an exponential moment, deduce that it has moments to all orders, i.e. that $\mathbb{E}(|X|^n) < \infty$ for all $n \in \mathbb{N}$.

ps:E_int **4.11** Let X be a real-valued random variable with law p_X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Show that for all bounded measurable functions $f : \mathbb{R} \to \mathbb{R}$,

$$\int_\Omega f(X(\omega))d\mathbb{P}(\omega) = \int_\mathbb{R} f(x)dp_X(x).$$

What can you say about these integrals when f is non-negative but not necessarily bounded?

Hint: Begin with f *an indicator function, then extend to simple, bounded non-negative and general bounded measurable functions.*

Chapter 5

Sequences of random variables

rv_sequences

In this section we think about sequences of random variables, and about taking limits of random variables. As with real numbers, sequences and limits are our main tool for justifying the use of approximations. Approximations allow us to better understand complicated models, by giving us a way to replace complicated random objects with simpler ones (whilst still maintaining some degree of accuracy). As such, this theory underpins much of stochastic modelling.

5.1 The Borel-Cantelli lemmas

orel-cantelli

The Borel-Cantelli lemmas are a tool for understanding the tail behaviour of a sequence (E_n) of events. The key definitions are

$$\begin{aligned}\{E_n \text{ i.o.}\} = \{E_n,\ \text{infinitely often}\} \quad &= \bigcap_m \bigcup_{n\geq m} E_n = \{\omega : \omega\in E_n \text{ for infinitely many } n\}\\ \{E_n \text{ e.v.}\} = \{E_n,\ \text{eventually}\} \quad &= \bigcup_m \bigcap_{n\geq m} E_n = \{\omega : \omega\in E_n \text{ for all sufficiently large } n\}.\end{aligned}$$

The set $\{E_n \text{ i.o.}\}$ is the event that infinitely many of the individual events E_n occur. The set $\{E_n \text{ e.v.}\}$ is the event that, for some (random) N, all the events E_n for which $n\geq N$ occur.

For example, we might take an infinite sequence of coin tosses and choose E_n to be the event that the n^{th} toss is a head. Then $\{E_n \text{ i.o.}\}$ is the event that infinitely many heads occur, and $\{E_n \text{ e.v.}\}$ is the event that, after some point, all remaining tosses show heads.

Note that by straightforward set algebra,

$$\Omega\setminus\{E_n \text{ i.o.}\} = \{\Omega\setminus E_n \text{ e.v.}\}. \tag{5.1}$$ {evio}

In our coin tossing example, $\Omega\setminus E_n$ is the event that the n^{th} toss is a tail. So (5.1) says that 'there are not infinitely many heads' if and only if 'eventually, we see only tails'.

The Borel-Cantelli lemmas, respectively, give conditions under which the probability of $\{E_n \text{ i.o.}\}$ is either 0 or 1.

lem:bc1 **Lemma 5.1.1 (First Borel-Cantelli Lemma)** *Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of events and suppose $\sum_{n=1}^\infty \mathbb{P}[E_n]<\infty$. Then $\mathbb{P}[E_n \text{ i.o.}]=0$.*

PROOF: We have

$$\mathbb{P}\left[\bigcap_N \bigcup_{n\geq N} E_n\right] = \lim_{N\to\infty}\mathbb{P}\left[\bigcup_{n\geq N} E_N\right] \leq \lim_{N\to\infty}\sum_{n=N}^\infty \mathbb{P}[E_n] = 0,$$

Here, the first step follows by applying Theorem 4.1.1 to the decreasing sequence of events (B_N) where $B_N = \bigcup_{n\geq N} E_n$. The second stop follows by Theorem 1.7.2 and the fact that limits preserve weak inequalities. The final step follows because $\sum_{n=1}^{\infty}\mathbb{P}[E_n] < \infty$. ■

For example, suppose that (X_n) are random variables that take the values 0 and 1, and that $\mathbb{P}[X_n = 1] = \frac{1}{n^2}$ for all n. Then $\sum_n \mathbb{P}[X_n = 1] = \sum_n \frac{1}{n^2} < \infty$ so, by Lemma 5.1.1, $\mathbb{P}[X_n = 1 \text{ i.o.}] = 0$, which by (5.1) means that $\mathbb{P}[X_n = 0 \text{ e.v.}] = 1$. So, almost surely, beyond some (randomly • located) point in our sequence (X_n), we will see only zeros. Note that we did not require the (X_n) to be independent.

lem:bc2 **Lemma 5.1.2 (Second Borel-Cantelli Lemma)** *Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of independent events and suppose that $\sum_{n=1}^{\infty}\mathbb{P}[E_n] = \infty$. Then $\mathbb{P}[E_n \text{ i.o.}] = 1$.*

PROOF: Write $E_n^c = \Omega\setminus E_n$. We will show that $\mathbb{P}[E_n^c \text{ e.v.}] = 0$, which by (5.1) implies our stated result. Note that

$$\mathbb{P}[E_n^c \text{ e.v.}] = \mathbb{P}\left[\bigcup_N \bigcap_{n\geq N} E_n^c\right] \leq \sum_{N=1}^{\infty}\mathbb{P}\left[\bigcap_{n\geq N} E_n^c\right] \tag{5.2}$$ {eq:Encleq}

by Theorem 1.7.2. Moreover, since the (E_n) are independent, so are the (E_n^c), so

$$\mathbb{P}\left[\bigcap_{n\geq N} E_n^c\right] = \prod_{n=N}^{\infty}\mathbb{P}[E_n^c] = \prod_{n=N}^{\infty}(1-\mathbb{P}[E_n]) \leq \prod_{n=N}^{\infty} e^{-\mathbb{P}[E_n]} = \exp\left(-\sum_{n=N}^{\infty}\mathbb{P}[E_n]\right) = 0.$$

Here, the first step follows by Exercise **4.5**. The second step is immediate and the third step uses that $1-x \leq e^{-x}$ for $x\in[0,1]$. The fourth step is immediate and the final step holds because $\sum_n \mathbb{P}[E_n] = \infty$. By (5.2) we thus have $\mathbb{P}[E_n^c \text{ e.v.}] = 0$. ■

For example, suppose that (X_n) are i.i.d. random variables such that $\mathbb{P}[X_n = 1] = \frac{1}{2}$ and $\mathbb{P}[X_n = -1] = \frac{1}{2}$. Then $\sum_n \mathbb{P}[X_n = 1] = \infty$ and, by Lemma 5.1.2, $\mathbb{P}[X_n = 1 \text{ i.o.}] = 1$. By symmetry, we have also $\mathbb{P}[X_n = 0 \text{ i.o.}] = 1$. So, if we look along our sequence, almost surely we will see infinitely many 1s and infinitely many 0s.

Since both the Borel-Cantelli lemmas come down to summing a series, a useful fact to remember from real analysis is that, for $p\in\mathbb{R}$,

$$\sum_{n=1}^{\infty} n^{-p} < \infty \quad\Leftrightarrow\quad p > 1.$$

Recall that this fact follows from the integral test for convergence of series.

5.2 Convergence of Random Variables

sec:conv

Let (X_n) be a sequence of random variables, all of which are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. There are various different ways in which we can examine the convergence of this sequence to a random variable X (which is also defined on $(\Omega, \mathcal{F}, \mathbb{P})$. They are called *modes of convergence*.

When we talk about convergence of real numbers $a_n \to a$ we only have one mode of convergence, which we might think of as convergence of the value of a_n to the value of a. Random variables are much more complicated objects; they take many different values with different probabilities. For this reason, there are multiple different modes of convergence of random variables.

We say that (X_n) converges to X

- *in distribution* if whenever $\mathbb{P}[X = x] = 0$ we have $\mathbb{P}[X_n \leq x] \to \mathbb{P}[X \leq x]$.
- *in probability* if given any $a > 0$, we have $\mathbb{P}[|X_n - X| > a] \to 0$ as $n \to \infty$,
- *almost surely* if $\mathbb{P}[X_n(\omega) \to X(\omega),\ \text{as } n \to \infty] = 1$,
- *in L^p* if $\mathbb{E}[|X_n - X|^p] \to 0$ as $n \to \infty$.

When (X_n) converges to X almost surely we sometimes write $X_n \to X$ a.s. as $n \to \infty$. We may also write the type of convergence above the arrow e.g. $X_n \stackrel{L^2}{\to} X$ or $X_n \stackrel{a.s.}{\to} X$.

For L^p convergence we are usually only interested in the cases $p = 1$ and $p = 2$. The case $p = 2$ is sometimes known as convergence in *mean square*. For reasons that are outside of the scope of this course, L^p convergence is only defined for $p \in [1, \infty)$.

Happily, there are some relationships between these different modes of convergence.

em:conv_modes

Lemma 5.2.1 *Let X_n, X be random variables.*

1. *If $X_n \stackrel{\mathbb{P}}{\to} X$ then $X_n \stackrel{d}{\to} X$.*
2. *If $X_n \stackrel{a.s.}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
3. *If $X_n \stackrel{L^p}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
4. *Let $1 \leq p < q$. If $X_n \stackrel{L^q}{\to} X$ then $X_n \stackrel{L^p}{\to} X$.*

In all other cases (i.e. that are not automatically implied by the above), convergence in one mode does not imply convergence in another.

PROOF: The last part follows from the examples in Exercises 5.1-5.3. We'll give proofs of parts 1-4 here.

Part 1. Let $x \in \mathbb{R}$ be such that $\mathbb{P}[X = x] = 0$ and let $\epsilon > 0$. We have

$$\begin{aligned} \mathbb{P}[X \leq x] &= \mathbb{P}[X \leq x, |X_n - X| < \epsilon] + \mathbb{P}[X \leq x, |X_n - X| \geq \epsilon] \\ &\leq \mathbb{P}[X_n \leq x + \epsilon] + \mathbb{P}[|X_n - X| \geq \epsilon]. \end{aligned} \tag{5.3}$$

{eq:pd1}

Putting $x - \epsilon$ in place of x we obtain

$$\mathbb{P}[X \leq x - \epsilon] \leq \mathbb{P}[X_n \leq x] + \mathbb{P}[|X_n - X \geq \epsilon]. \tag{5.4}$$

{eq:pd2}

Combining (5.3) and (5.4) leads to

$$\mathbb{P}[X \le x-\epsilon] - \mathbb{P}[|X_n - X| \ge \epsilon] \;\le\; \mathbb{P}[X_n \le x] \;\le\; \mathbb{P}[X \le x+\epsilon] + \mathbb{P}[|X_n - X| \ge \epsilon].$$

By Remark 4.2.2 and the fact that $\mathbb{P}[X = x] = 0$ we have that $y \mapsto \mathbb{P}[X \le y]$ is continuous at $y = x$. Hence, letting $\epsilon \to 0$, both $\mathbb{P}[X \le x+\epsilon]$ and $\mathbb{P}[X \le x-\epsilon]$ converge to $\mathbb{P}[X \le x]$. Since limits preserve weak inequalities this gives

$$\mathbb{P}[X \le x] - \mathbb{P}[|X_n - X| \ge \epsilon] \;\le\; \mathbb{P}[X_n \le x] \;\le\; \mathbb{P}[X \le x] + \mathbb{P}[|X_n - X| \ge \epsilon].$$

Letting $n \to \infty$ and using that $X_n \xrightarrow{\mathbb{P}} X$, the sandwich rule gives $\mathbb{P}[X_n \le x] \to \mathbb{P}[X \le x]$.

Part 2. Let $\epsilon > 0$ be arbitrary and let $A_n = \bigcup_{m=n}^\infty \{|X_m - X| > \epsilon\}$. Then (A_n) is a decreasing sequence of events. Let $A = \bigcap_{n=1}^\infty A_n$. If $\omega \in A$ then $X_n(\omega)$ cannot converge to $X(\omega)$ as $n \to \infty$ and so $\mathbb{P}[A] = 0$, because $X_n \to X$ almost surely. By Theorem 4.1.1 part (2), $\lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A] = 0$. But then by monotonicity,

$$\mathbb{P}[|X_n - X| > \epsilon] \le \mathbb{P}[A_n] \to 0 \text{ as } n \to \infty.$$

Part 4. $(\star)$ We'll prove part 4 next, because it will be helpful in part 3 below. This part is non-examinable because we need an inequality that is (just) outside of our own syllabus. It is true that

$$\mathbb{E}[|X|]^p \le \mathbb{E}[|X|^p] \tag{5.5}$$ {eq:jensen_p

for all $p \ge 1$ and all random variables X. For general p, this is a special case of a result known as *Jensen's inequality*, which we don't cover in this course. However, the $p = 2$ case is a consequence of the Cauchy-Schwartz inequality, which appears in these notes as Exercise **4.9**. We have already seen the $p = 1$ case in equation (3.7) and Exercise **3.9**.

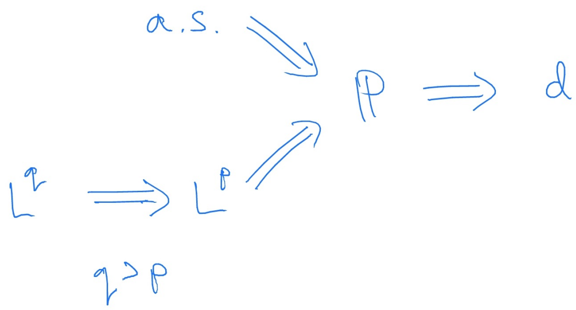
Putting $|X_n - X|^p$ into (5.5), and then putting $q/p \ge 1$ in place of p, we obtain that $\mathbb{E}[|X_n - X|^p]^{q/p} \le (\mathbb{E}[|X_n - X|^{p(q/p)}])$. Thus $\mathbb{E}[|X_n - X|^p] \le (\mathbb{E}[|X_n - X|^q])^{p/q}$. The result follows.

Part 3. Thanks to part 4, it suffices to prove that L^1 convergence implies convergence in probability. From Markov's inequality (Lemma 3.3.2) for any $a > 0$ we have

$$\mathbb{P}[|X_n - X| > a] \le \frac{\mathbb{E}[|X_n - X|]}{a}.$$

If $X_n \xrightarrow{L^1} X$ then the right hand side tends to zero as $n \to \infty$, hence so does the left. ■

Remark 5.2.2 *The following diagram records which modes of convergence imply which other modes of convergence.*



Recall that for real numbers, if $a_n \to a$ and $a_n \to b$ then $a = b$, which is known as uniqueness of limits. For random variables, the situation is a little more complicated: if $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} Y$ then $X = Y$ almost surely. By Lemma 5.2.1, this result also applies to $\stackrel{L^p}{\to}$ and $\stackrel{a.s.}{\to}$. However, if we have only $X_n \stackrel{d}{\to} X$ and $X_n \stackrel{d}{\to} Y$ then we can only conclude that X and Y have the same distribution, that is $\mathbb{P}[X \leq x] = \mathbb{P}[Y \leq x]$ for all x. Proving these facts is Exercise 5.7.

Establishing convergence in distribution, probability and L^p usually comes down to calculating (or estimating) the important quantities involved: $\mathbb{P}[X \leq x]$, $\mathbb{P}[|X_n - X| \leq a]$ and $\mathbb{E}[|X_n - X|^p]$, and then thinking about their limits as $n \to \infty$. There are several examples of this type within the exercises. Almost sure convergence is harder to work with, but here we can often use the Borel-Cantelli lemmas.

Example 5.2.3 Let (X_n) be a sequence of i.i.d. random variables, each with the uniform distribution on $[0,1]$. Then $\mathbb{P}[X_n \leq \frac{1}{3}] = \mathbb{P}[X_n \geq \frac{2}{3}] = \frac{1}{3}$, so (using independence) by two applications of the second Borel-Cantelli lemma we have $\mathbb{P}[X_n \leq \frac{1}{3}\text{ i.o.}] = \mathbb{P}[X_n \geq \frac{2}{3}\text{ i.o.}] = 1$. Hence, with probability 1, the sequence X_n will oscillate infinitely often between $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$, in which case it cannot converge. Thus (X_n) does not converge almost surely (to any limit) in this case.

Alternatively, consider a sequence (Y_n) with $\mathbb{P}[Y_n = 1] = \frac{1}{n^2}$ and $\mathbb{P}[Y_n = 0] = 1 - \frac{1}{n^2}$. Since $\sum \frac{1}{n^2} < \infty$, the first Borel-Cantelli lemma tells us that $\mathbb{P}[Y_n = 1\text{ i.o.}] = 0$. Hence $\mathbb{P}[Y_n = 0\text{ e.v.}] = 1$, which implies that $\mathbb{P}[Y_n \to 0] = 1$, or in other words that $Y_n \stackrel{a.s.}{\to} 0$.

lem:sub **Lemma 5.2.4** *If $X_n \to X$ in probability as $n \to \infty$ then there is a subsequence of (X_n) that converges to X almost surely.*

PROOF: If (X_n) converges in probability to X, for all $c > 0$, given any $\epsilon > 0$, there exists $N(c) \in \mathbb{N}$ so that for all $n \geq N(c)$, $\mathbb{P}[|X_n - X| > c] < \epsilon$.

In order to find our subsequence:

– First choose, $c = 1$ and $\epsilon = 1/2$, then for $n \geq N(1)$, $\mathbb{P}[|X_n - X| > 1] < 1/2$.

– Next choose $c = 1/2$ and $\epsilon = 1/4$, then for $n \geq N(2)$, $\mathbb{P}[|X_n - X| > 1/2] < 1/4$.

– In general, choose $c = 1/r$ and $\epsilon = 1/2^r$, then for $n \geq N(r)$, $\mathbb{P}[|X_n - X| > 1/r] < 1/2^r$.

Set $k_r = \max\{N(1), N(2), \ldots, N(r), r\}$ for $r \in \mathbb{N}$. to obtain a subsequence (X_{k_r}) so that for all $r \in \mathbb{N}$,

$$\mathbb{P}[|X_{k_r} - X| > 1/r] < 1/2^r.$$

Since $\sum \frac{1}{2^r} < \infty$, by the first Borel-Cantelli lemma (Lemma 5.1.1) we have

$$\mathbb{P}[|X_{k_r} - X| > 1/r \ \ i.o.] = 0,$$

and so

$$\mathbb{P}[|X_{k_r} - X| \leq 1/r \ \ e.v.] = 1.$$

Hence, almost surely, there exists some $R \in \mathbb{N}$ such that for all $r \geq R$ we have $|X_{k_r} - X| < 1/r$, which implies that $|X_{k_r} - X| \to 0$. Hence $X_{k_r} \to X$ almost surely. ■

5.3 Laws of Large Numbers

Let (X_n) be a sequence of random variables all defined on the same probability space, that have the following properties,

- they are independent;
- they are identically distributed i.e. $p_{X_n} = p_{X_m}$ for all $n \neq m$. In other words, for all $A \in \mathcal{B}(\mathbb{R})$,
$$\mathbb{P}[X_1 \in A] = \mathbb{P}[X_2 \in A] = \cdots = \mathbb{P}[X_n \in A] = \cdots$$

Such a sequence is said to be *'i.i.d.'*. Sequences of this type are very important in modelling (consider the steps of a random walk) and also statistics (consider a sequence of idealised experiments carried out under identical conditions).

We are interested in the *empirical mean*

$$\overline{X_n} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

which in statistics is often known as the sample mean. If X_n is integrable for some (and hence all) $n \in \mathbb{N}$, with $\mu = \mathbb{E}[X_n]$, then by linearity $\overline{X_n}$ is integrable and $\mathbb{E}[\overline{X_n}] = \mu$. It is extremely important to learn about the asymptotic behaviour of $\overline{X_n}$ as $n \to \infty$. Two key results are the *weak law of large numbers* (WLLN) and the *strong law of large numbers* (SLLN). The strong law implies the weak law, but it is much trickier to prove. We will include a full proof of the weak law and a partial proof of the strong law.

thm:wlln **Theorem 5.3.1 (WLLN)** *Let (X_n) be a sequence of integrable i.i.d. random variables with $\mathbb{E}(X_n) = \mu$ for all $n \in \mathbb{N}$. Suppose also that $\mathbb{E}[X_n^2] < \infty$ for all $n \in \mathbb{N}$. Then $\overline{X_n} \to \mu$ in probability as $n \to \infty$.*

PROOF: Let $\sigma^2 = \mathrm{var}(X_n)$ for all $n \in \mathbb{N}$. Since the (X_n) are independent, it follows that $\mathrm{var}(\overline{X_n}) = \frac{\sigma^2}{n}$. Then by Chebychev's inequality, for all $a > 0$,

$$\mathbb{P}\left[|\overline{X_n} - \mu| > a\right] \leq \frac{\mathrm{var}(\overline{X_n})}{a^2} = \frac{\sigma^2}{na^2}$$

which tends to zero as $n \to \infty$. ∎

thm:slln **Theorem 5.3.2 (SLLN)** *Let (X_n) be a sequence of integrable i.i.d. random variables with $\mathbb{E}(X_n) = \mu$ for all $n \in \mathbb{N}$. Then $\overline{X_n} \to \mu$ almost surely as $n \to \infty$.*

The full proof of the SLLN is too long for this course. You can find it in e.g. Rosenthal on pages 47-49. We'll give the proof in a special case, by making an assumption on the fourth moments of the sequence (X_n), and then remark on the general case.

ass:4slln **Assumption 5.3.3** $\mathbb{E}[(X_n - \mu)^4] = b < \infty$ for all $n \in \mathbb{N}$.

PROOF: [Of Theorem 5.3.2 under Assumption 5.3.3.] Without loss of generality we may assume that $\mu = 0$. To see this, instead of X_n we would consider $Y_n = X_n - \mu$.

Let $S_n = X_1 + X_2 + \cdots + X_n$ so that $S_n = n\overline{X_n}$ for all $n \in \mathbb{N}$. Consider $\mathbb{E}[S_n^4]$. It contains many terms of the form $\mathbb{E}[X_j X_k X_l X_m]$ with distinct indices and these all vanish by independence, because $\mathbb{E}[X_j] = \mathbb{E}[X_k] = \mathbb{E}[X_l] = \mathbb{E}[X_m] = 0$. A similar argument disposes of terms of the form $\mathbb{E}[X_j X_k^3]$ and $\mathbb{E}[X_j X_k X_l^2]$, where j, k, l are distinct. The only terms with non-vanishing expectation are n terms of the form X_i^4 and $\binom{n}{2}\binom{4}{2} = 3n(n-1)$ terms of the form $X_i^2 X_j^2$ with $i \neq j$. By Theorem 4.4.2, X_i^2 and X_j^2 are independent for $i \neq j$ and so by Theorem 4.4.2

$$\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]E[X_j^2] = \mathrm{var}(X_i^2)\,\mathrm{var}(X_j^2) = \sigma^4.$$

Putting all this together,

$$\begin{aligned}\mathbb{E}[S_n^4] &= \sum_{i=1}^{n} \mathbb{E}[X_i^4] + \sum_{i \neq j} \mathbb{E}[X_i^2 X_j^2] \\ &= nb + 3n(n-1)\sigma^4 \leq Kn^2,\end{aligned}$$

where $K = b + 3\sigma^4$. For all $a > 0$, by Markov's inequality (Lemma 3.3.1)

$$\mathbb{P}[|\overline{X_n}| > a] = \mathbb{P}[S_n^4 > a^4 n^4] \leq \frac{\mathbb{E}[S_n^4]}{a^4 n^4} \leq \frac{Kn^2}{a^4 n^4} = \frac{K}{a^4 n^2}.$$

Recall that $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. Hence, by the first Borel-Cantelli lemma, $\mathbb{P}[|\overline{X_n}| > a \text{ i.o.}] = 0$ and so $\mathbb{P}[|\overline{X_n}| \leq a \text{ e.v.}] = 1$. It follows that $\overline{X_n} \to 0$ a.s. as required. ■

Remark 5.3.4 *($\star$) The proof, in the general case without Assumption 4.1, uses a truncation argument, based on $Y_n = X_n \mathbb{1}_{\{X_n \leq n\}}$. Then $Y_n \leq n$ for all n and so $\mathbb{E}[Y_n^k] \leq n^k$ for all k. If $X_n \geq 0$ for all n, $\mathbb{E}[Y_n] \to \mu$ by monotone convergence. Roughly speaking the argument is to prove a SLLN for the $\overline{Y_n}$s, and then transfer this to the $\overline{X_n}$s. It requires much more work.*

If $\mathbb{E}(X_n^2) < \infty$ then (by Problem **4.9**) $\mathrm{var}(X_n) = \sigma^2 < \infty$ for some (and hence all) $n \in \mathbb{N}$. It follows by elementary properties of the variance that $\mathrm{var}(\overline{X_n}) = \frac{\sigma^2}{n}$. This is known as the empirical variance or sample variance. The Central Limit Theorem investigates this quantity, as we will see in Section 5.5.

5.4 Characteristic Functions and Weak Convergence ($\star$)

In this section, we introduce two tools that we will need to prove the central limit theorem. They rely on Lebesgue integration in $\mathbb{C}$, which we introduced in Remark 3.6.4.

5.4.1 Characteristic Functions ($\star$)

Let (S, Σ, m) be a measure space and $f : S \to \mathbb{C}$ be a complex-valued function. Then we can write $f = f_1 + if_2$ where f_1 and f_2 are real-valued functions. We say that f is measurable/integrable if both f_1 and f_2 are. Define $|f|(x) = |f(x)| = \sqrt{f_1(x)^2 + f_2(x)^2}$ for each $x \in S$. It is not difficult to see that $|f|$ is measurable, using e.g. Problem **2.4**. In Problem **5.8**, you can prove that f is integrable if and only if $|f|$ is integrable. The Lebesgue dominated convergence theorem continues to hold for sequences of measurable functions from S to $\mathbb{C}$.

Now let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Its *characteristic function* $\phi_X : \mathbb{R} \to \mathbb{C}$ and is defined, for each $u \in \mathbb{R}$, by

$$\phi_X(u) = \mathbb{E}(e^{iuX}) = \int_{\mathbb{R}} e^{iuy} p_X(dy).$$

Note that $y \to e^{iuy}$ is measurable since $e^{iuy} = \cos(uy) + i\sin(uy)$ and integrability holds since $|e^{iuy}| \leq 1$ for all $y \in \mathbb{R}$ and in fact we have $|\phi_X(u)| \leq 1$ for all $u \in \mathbb{R}$.

Example $X \sim N(\mu, \sigma^2)$ means that X has a normal or Gaussian distribution with mean μ and variance σ^2 so that for all $x \in \mathbb{R}$,

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} dy.$$

In Problem **5.8**, you can show for yourself that in this case, for all $u \in \mathbb{R}$

$$\phi_X(u) = \exp\left\{i\mu u - \frac{1}{2}\sigma^2 u^2\right\}.$$

Characteristic functions have many interesting properties. Here is one of the most useful. It is another instance of the “independence means multiply” philosophy.

thm:cfind **Theorem 5.4.1** *If X and Y are independent random variables then for all $u \in \mathbb{R}$,*

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u).$$

PROOF:

$$\phi_{X+Y}(u) = \mathbb{E}(e^{iu(X+Y)}) = \mathbb{E}(e^{iuX}e^{iuY}) = \mathbb{E}(e^{iuX})\mathbb{E}(e^{iuY}) = \phi_X(u)\phi_Y(u),$$

by Theorem **??**. ■

The following result is also important but we omit the proof. It tells us that the probability law of a random variable is uniquely determined by its characteristic function.

Theorem 5.4.2 *If X and Y are two random variables for which $\phi_X(u) = \phi_Y(u)$ for all $u \in \mathbb{R}$ then $p_X = p_Y$.*

The characteristic function is the Fourier transform of the law p_X of the random variable X and we have seen that it always exists. In elementary probability theory courses we often meet the Laplace transform $\mathbb{E}(e^{uX})$ of X which is called the *moment generating function*. This exists in some nice cases (e.g. when X is Gaussian), but will not do so in general as $y \to e^{uy}$ may not be integrable since it becomes unbounded as $y \to \infty$ (when $u > 0$) and as $y \to -\infty$ (when $u < 0$.)

We will now develop an important inequality that we will need to prove the central limit theorem. Let $x \in \mathbb{R}$ and let $R_n(x)$ be the remainder term of the series expansion at $n \in \mathbb{N} \cup \{0\}$ in e^{ix}, i.e.

$$R_n(x) = e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}.$$

Note that $R_0(x) = e^{ix} - 1 = \begin{cases} \int_0^x ie^{iy}dy & \text{if } x > 0 \\ -\int_x^0 ie^{iy}dy & \text{if } x < 0 \end{cases}$. From the last two identities, we have $|R_0(x)| \leq \min\{|x|, 2\}$. Then you should check that $R_n(x) = \begin{cases} \int_0^x iR_{n-1}(y)dy & \text{if } x > 0 \\ -\int_x^0 iR_{n-1}(y)dy & \text{if } x < 0 \end{cases}$. Finally using induction, we can deduce the useful inequality:

$$|R_n(x)| \leq \min\left\{\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\right\}.$$

Now let X be a random variable with characteristic function ϕ_X for which $\mathbb{E}(|X|^n) < \infty$ for some given $n \in \mathbb{N}$. Then integrating the last inequality yields for all $y \in \mathbb{R}$

$$\left|\phi_X(y) - \sum_{k=0}^{n} \frac{(iy)^k \mathbb{E}(X^k)}{k!}\right| \leq \mathbb{E}\left[\min\left\{\frac{2|yX|^n}{n!}, \frac{|yX|^{n+1}}{(n+1)!}\right\}\right]. \tag{5.6}$$

{remain}

When we prove the CLT we will want to apply this in the case $n = 2$ to a random variable that has $\mathbb{E}(X) = 0$. Then writing $\mathbb{E}(X^2) = \sigma^2$ we deduce that for all $u \in \mathbb{R}$.

$$\left|\phi_X(y) - 1 + \frac{1}{2}\sigma^2 y^2\right| \leq \theta(y), \tag{5.7}$$

{remain2}

where $\theta(y) = y^2 \mathbb{E}\left[\min\left\{|X|^2, |y|\frac{|X|^3}{6}\right\}\right]$. Note that

$$\min\left\{|X|^2, |y|\frac{|X|^3}{6}\right\} \leq |X|^2$$

which is integrable by assumption. Also we have

$$\lim_{y \to 0} \min\left\{|X|^2, |y|\frac{|X|^3}{6}\right\} = 0,$$

and so by the dominated convergence theorem we can deduce the following important property of θ which is that

$$\lim_{y \to 0} \frac{\theta(y)}{y^2} = 0. \tag{5.8}$$

{remain3}

5.4.2 Weak Convergence ($\star$)

A sequence (μ_n) of probability measures on $\mathbb{R}$ is said to converge *weakly* to a probability measure μ if

$$\lim_{n\to\infty}\int_{\mathbb{R}} f(x)\mu_n(dx) = \int_{\mathbb{R}} f(x)\mu(dx)$$

for all bounded continuous functions f defined on $\mathbb{R}$. In the case where there is a sequence of random variables (X_n) and instead of μ_n we have p_{X_n} and also μ is the law p_X of a random variable X we say that (X_n) *converges in distribution* to X; so that convergence in distribution means the same thing as weak convergence of the sequence of laws. It can be shown that (X_n) converges to X in distribution if and only if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at every continuity point x of the c.d.f. F, which matches the definition we gave in Section 5.2.

There is an important link between the concepts of weak convergence and characteristic functions which we present next.

thm:LCT **Theorem 5.4.3** *Let (X_n) be a sequence of random variables, where each X_n has characteristic function ϕ_n and let X be a random variable having characteristic function ϕ. Then (X_n) converges to X in distribution if and only if $\lim_{n\to\infty}\phi_n(u) = \phi(u)$ for all $u \in \mathbb{R}$.*

PROOF: We'll only do the easy part here. Suppose (X_n) converges to X in distribution. Then

$$\begin{aligned}\phi_n(u) &= \int_{\mathbb{R}} \cos(uy)p_{X_n}(dy) + i\int_{\mathbb{R}} \sin(uy)p_{X_n}(dy)\\ &\to \int_{\mathbb{R}} \cos(uy)p_X(dy) + i\int_{\mathbb{R}} \sin(uy)p_X(dy) = \phi(u),\end{aligned}$$

as $n\to\infty$, since both $y \to \cos(uy)$ and $y \to \sin(uy)$ are bounded continuous functions. See e.g. Rosenthal pp.108-9 for the converse [1]. ■

[1] This reference is to the first edition. You'll find it on pp. 132-3 in the second edition

5.5 The Central Limit Theorem ($\star$)

sec:clt

Let (X_n) be a sequence of i.i.d. random variables having finite mean μ and finite variance σ^2. We have already met the SLLN which tells us that $\overline{X_n}$ converges to μ a.s. as $n \to \infty$. Note that the standard deviation (i.e. the square root of the variance) of $\overline{X_n}$ is $\sigma/\sqrt{n}$ which also converges to zero as $n \to \infty$. Now consider the sequence (Y_n) of standardised random variables defined by

$$Y_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \tag{5.9}$$ {stand}

Then $\mathbb{E}(Y_n) = 0$ and $\mathrm{Var}(Y_n) = 1$ for all $n \in \mathbb{N}$.

Its difficult to underestimate the importance of the next result. It shows that the normal distribution has a universal character as the attractor of the sequence (Y_n). From a modelling point of view, it tells us that as you combine together many i.i.d. different observations then they aggregate to give a normal distribution. This is of vital importance in applied probability and statistics. Note however that if we drop our standing assumption that all the X_n's have a finite variance, then this would no longer be true.

thm:CLT **Theorem 5.5.1 (Central Limit Theorem)** *Let (X_n) be a sequence of i.i.d. random variables each having finite mean μ and finite variance σ^2. Then the corresponding sequence (Y_n) of standardised random variables converges in distribution to the standard normal $Z \sim N(0,1)$, i.e. for all $a \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}(Y_n \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}y^2}\,dy.$$

Before we give the proof, we state a known fact from elementary analysis. We know that for all $y \in \mathbb{R}$,

$$\lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n = e^y.$$

Now for all $y \in \mathbb{R}$, let $(\alpha_n(y))$ be a sequence of real (or complex) numbers for which $\lim_{n\to\infty} \alpha_n(y) = 0$. Then we also have that for all $y \in \mathbb{R}$

$$\lim_{n\to\infty} \left(1 + \frac{y + \alpha_n(y)}{n}\right)^n = e^y \tag{5.10}$$ {anal}

You may want to try your hand at proving this rigorously.

PROOF: For convenience we assume that $\mu = 0$ and $\sigma = 1$. Indeed if it isn't we can just replace X_n everywhere by $(X_n - \mu)/\sigma$. Let ψ be the common characteristic function of the X_ns so that in particular $\psi(u) = \mathbb{E}(e^{iuX_1})$ for all $u \in \mathbb{R}$. Let ϕ_n be the characteristic function of Y_n for each $n \in \mathbb{N}$. Then for all $u \in \mathbb{R}$, using Theorem 5.4.1 we find that

$$\begin{aligned}
\phi_n(u) &= \mathbb{E}(e^{iuS_n/\sqrt{n}}) \\
&= \mathbb{E}\left(e^{iu(\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)}\right) \\
&= \psi(u/\sqrt{n})^n \\
&= \mathbb{E}\left(e^{i\frac{u}{\sqrt{n}}X_1}\right)^n \\
&= \left(1 + \frac{iu}{\sqrt{n}}\mathbb{E}(X_1) - \frac{u^2}{2n}\mathbb{E}(X_1^2) + \frac{\theta_n(u)}{n}\right)^n,
\end{aligned}$$

where by (5.7) and the same argument we used to derive (5.8),

$$|\theta_n(u)| \leq u^2 \mathbb{E}\left[\min\left\{|X_1|^2, \frac{|u|.|X_1|^3}{6\sqrt{n}}\right\}\right] \to 0$$

as $n \to \infty$, for all $u \in \mathbb{R}$.

Now we use (5.10) to find that

$$\begin{aligned}\phi_n(u) &= \left(1 - \frac{\frac{u^2}{2} - \theta_n(u)}{n}\right)^n \\ &\to e^{-\frac{1}{2}u^2} \text{ as } n \to \infty.\end{aligned}$$

The result then follows by Theorem 5.4.3. ■

5.5.1 Further discussion ($\star$)

The CLT may be extensively generalised. We mention just a few such results here. If the i.i.d. sequence (X_n) is such that $\mu = 0$ and $\mathbb{E}(|X_n|^3) = \rho^3 < \infty$, the *Berry-Esseen theorem* gives a useful bound for the difference between the cdf of the normalised sum and the cdf Φ of the standard normal. To be precise we have that for all $x \in \mathbb{R}, n \in \mathbb{N}$:

$$\left|\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x)\right| \leq C\frac{\rho}{\sqrt{n}\sigma^3},$$

where $C > 0$.

We can also relax the requirement that the sequence (X_n) be i.i.d.. Consider the *triangular array* $(X_{nk}, k = 1, \ldots, n, n \in \mathbb{N})$ of random variables which we may list as follows:

$$\begin{array}{lllll} X_{11} & & & & \\ X_{21} & X_{22} & & & \\ X_{31} & X_{32} & X_{33} & & \\ \vdots & \vdots & \vdots & & \\ X_{n1} & X_{n2} & X_{n3} & \ldots & X_{nn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

We assume that each row comprises independent random variables. Assume further that $\mathbb{E}(X_{nk}) = 0$ and $\sigma_{nk}^2 = \mathbb{E}(X_{nk}^2) < \infty$ for all k, n. Define the row sums $S_n = X_{n1} + X_{n2} + \cdots + X_{nn}$ for all $n \in \mathbb{N}$ and define $\tau_n = \mathrm{var}(S_n) = \sum_{k=1}^n \sigma_{nk}^2$. *Lindeburgh's central limit theorem* states that if we have the asymptotic tail condition

$$\lim_{n\to\infty} \sum_{k=1}^n \frac{1}{\tau_n^2} \int_{|X_{nk}| \geq \epsilon\tau_n} X_{nk}^2(\omega)d\mathbb{P}(\omega) = 0,$$

for all $\epsilon > 0$ then $\frac{S_n}{\tau_n}$ converges in distribution to a standard normal as $n \to \infty$.

The highlights of this last chapter have been the proofs of the law of large numbers and central limit theorem. There is a third result that is often grouped together with the other two as one of the key results about sums of i.i.d. random variables. It is called the *law of the iterated logarithm* and it gives bounds on the fluctuations of S_n for an i.i.d sequence with $\mu = 0$ and $\sigma = 1$. The result is quite remarkable. It states that

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log(n)}} = 1 \text{ a.s.} \tag{5.11}$$ {lil}

This means that (with probability one) if $c > 1$ then only finitely many of the events $S_n > c\sqrt{2n\log\log(n)}$ occur but if $c < 1$ then infinitely many of such events occur. This gives a *very* precise description of the long-term behaviour of S_n.

You should be able to deduce from (5.11) that

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log(n)}} = -1 \text{ a.s.}$$

5.6 Exercises

ps:D_not_P

5.1 Let (X_n) be a sequence of i.i.d. random variables such that $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = 0] = \frac{1}{2}$. Show that $X_n \stackrel{d}{\to} X_1$ as $n \to \infty$, but that this convergence does not hold in probability.

Hint: Use one of the Borel-Cantelli lemmas.

rel_cantelli

5.2 (a) Let (X_n) be a sequence of random variables such that $\mathbb{P}[X_n = n] = \frac{1}{n^2}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n^2}$. Show that $X_n \stackrel{a.s}{\to} 0$ and $X_n \stackrel{L^1}{\to} 0$.

(b) Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = n] = \frac{1}{n}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$. Show that X_n does not converge to zero almost surely or in L^1.

(c) Let (X_n) be a sequence of random variables such that $\mathbb{P}[X_n = n^2] = \frac{1}{n^2}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n^2}$. Show that $X_n \stackrel{a.s.}{\to} 0$, and that X_n does not converge to zero in L^1.

(d) Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = \sqrt{n}] = \frac{1}{n}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$. Show that $X_n \stackrel{L^1}{\to} 0$, and that X_n does not converge to almost surely to zero.

(e) Deduce that $X_n \stackrel{\mathbb{P}}{\to} 0$ in all of the above cases.

ps:P_not_as

5.3 Show that the following sequence (X_n) and candidate limit X of random variables converges in probability but not almost surely.

Take $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and $\mathbb{P}$ to be Lebesgue measure. Take $X = 0$ and define $X_n = \mathbb{1}_{A_n}$ where $A_1 = [0,1/2], A_2 = [1/2,1], A_3 = [0,1/4], A_4 = [1/4,1/2], A_5 = [1/2,3/4], A_6 = [3/4,1], A_7 = [0,1/8], A_8 = [1/8,1/4]$ etc.

ep_coin_runs

5.4 Let $k \in \mathbb{N}$. Prove that in a sequence of independent coin tosses, infinitely many runs of k consecutive heads will occur.

ps:ev_io

5.5 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let (A_n) be a sequence of events.

(a) Show that $\{A_n \text{ e.v.}\} \subseteq \{A_n \text{ i.o.}\}$.

(b) Show that $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ e.v.}\}$ and deduce that $\mathbb{P}[A_n \text{ i.o.}] = 1 - \mathbb{P}[A_n^c \text{ e.v.}]$.

(c) Show that

$$\mathbb{P}[A_n \text{ e.v.}] \le \liminf_{n\to\infty} \mathbb{P}[A_n] \le \limsup_{n\to\infty} \mathbb{P}[A_n] \le \mathbb{P}\left[A_n \text{ i.o.}\right].$$

s:wlln_uncor

5.6 Show that the conclusion of the weak law of large numbers continues to hold if the requirement that the random variables (X_n) are i.i.d. is replaced by the weaker condition that they are identically distributed, and *uncorrelated*, i.e. $\mathbb{E}[X_m X_n] = \mathbb{E}[X_m]\mathbb{E}[X_n]$ whenever $m \neq n$.

For this question, you should find an argument that doesn't rely on the strong law!

unique_limit

5.7 Let (X_n) be a sequence of random variables, and let X and Y be random variables.

(a) Show that if $X_n \stackrel{d}{\to} X$ and $X_n \stackrel{d}{\to} Y$ then X and Y have the same distribution.

(b) Show that if $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} Y$ then $X = Y$ almost surely.

ps:C_int_abs

5.8 ($\star$) Let (S, Σ) be a measurable space and $f : S \to \mathbb{C}$ be a (complex-valued) measurable function. Deduce that f is integrable if and only if $|f|$ is.

:N_char_func **5.9** ($\star$) If X is a normally distributed random variable with mean μ and variance σ^2, deduce that its characteristic ϕ_X function is given for each $u \in \mathbb{R}$ by $\phi_X(u) = \exp\left\{i\mu u - \frac{1}{2}\sigma^2 u^2\right\}$.

Hint: First show that it is sufficient to establish the case $\mu = 0$ and $\sigma = 1$ by writing $Y = \frac{1}{\sigma}(X - \mu)$. Then show that $y \to \phi_Y(u)$ is differentiable and deduce that $\phi_Y'(u) = -u\phi_Y(u)$. Now solve the initial value problem using what you know about $\phi_Y(0)$.

:gen_moments **5.10** ($\star$) Suppose that X is a random variable for which $\mathbb{E}(|X|^n) < \infty$ for some n. Explain carefully why

$$\mathbb{E}(X^n) = i^{-n} \left.\frac{d^n}{du^n}\phi_X(u)\right|_{u=0}.$$

lt_bernoulli **5.11** ($\star$) The first central limit theorem (CLT) to be established was due to de Moivre and Laplace. In this case each X_n takes only two values, 1 with probability p and 0 with probability $1-p$ where $0 < p < 1$ (i.e. the X_ns are i.i.d. *Bernoulli* random variables.) Write down the form of the CLT in this case (writing the standardised random variable Y_n in terms of $S_n = X_1 + X_2 + \cdots + X_n$), and explain its relation to the "binomial approximation to the normal distribution".

Chapter 6

Product Measures and Fubini's Theorem (Δ)

product_meas

Note that this chapter is marked with a (Δ). This means it is *only* included in MAS452/6352. It is not included in MAS350. It extends Sections 1.8 and 3.9.1.

In this chapter, we give brief notes which are intended as a summary of additional reading that you are expected to do outside the lectures. They emphasise the main ideas and concepts, but you will need to work carefully through all the relevant proofs. The recommended source is Adams and Guillemin, section 2.5, pp.89-102. *This material is examinable.* It can be studied straight after Chapter 3.

The aim of this section is to learn more about product measures, and how the theory of Lebesgue integration deals with multiple (double, triple, etc.) integrals. Before delving further into the details of these ideas, we'll need an additional tool from the theory of σ-fields.

6.1 Dynkin's $\pi - \lambda$ Lemma (Δ)

Let (S, Σ) be a measurable space. A collection $\mathcal{P}$ of sets in S is called a *π-system* if $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$ (i.e. $\mathcal{P}$ is closed under intersections).

A collection $\mathcal{L}$ of sets in S is called a *λ-system* if

(L1) $S \in \mathcal{L}$.

(L2) If (E_n) is an increasing sequence of sets in $\mathcal{L}$ (so $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$), then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{L}$.

(L3) If $E, F \in \mathcal{L}$ and $F \subset E$ then $E - F \in \mathcal{L}$.

Note that by (L1) and (L3), λ-systems are closed under complements.

p:ls_systems

Proposition 6.1.1 *If $\mathcal{L}$ is a λ-system that is also a π-system, then it is a σ-field.*

Proof: Since $\mathcal{L}$ is closed under complements and finite intersections, it is closed under finite unions by de Morgan's laws. To show that $\mathcal{L}$ is a σ-algeba, we need to prove that if (A_n) is an arbitrary sequence of sets in $\mathcal{L}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{L}$. This follows by writing $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ and using (L2), where

$$B_1 = A_1, B_2 = B_1 \cup (A_2 - B_1), B_3 = B_2 \cup (A_3 - B_2) \ldots$$

■

Recall that if $\mathcal{A}$ is a collections of sets in S, then $\sigma(\mathcal{A})$ is the smallest σ-field which contains $\mathcal{A}$.

Lemma 6.1.2 (Dynkin's $\pi - \lambda$ Lemma) *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system with $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

PROOF: It suffices to prove that $\mathcal{L}(\mathcal{P})$, which is the smallest λ-system that contains $\mathcal{P}$, is a σ-field. By Proposition 6.1.1, its enough to prove that $\mathcal{L}(\mathcal{P})$ is closed under intersections.

Step 1. Fix $A \in \mathcal{L}(\mathcal{P})$ and define

$$\mathcal{G}_A = \{B \subseteq \Omega; A \cap B \in \mathcal{L}(\mathcal{P})\}.$$

You should check that $\mathcal{G}_A$ is a λ-system.

Step 2. If $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P} \subseteq \mathcal{L}(\mathcal{P})$. Hence $B \in \mathcal{G}_A$. So we've shown that $\mathcal{P} \subseteq \mathcal{G}_A$, when $A \in \mathcal{P}$. But in Step 1, we proved that $\mathcal{G}_A$ is a λ-system, and so since $\mathcal{L}(\mathcal{P})$ is the smallest such, we deduce that $\mathcal{L}(\mathcal{P}) \subseteq \mathcal{G}_A$. This means in particular that if $A \in \mathcal{P}, B \in \mathcal{L}(\mathcal{P})$, then $A \cap B \in \mathcal{L}(\mathcal{P})$.

Step 3. If $A \in \mathcal{L}(\mathcal{P})$, then by Step 2 we have $A \cap B = B \cap A \in \mathcal{L}(\mathcal{P})$ when $B \in \mathcal{P}$. This shows that $\mathcal{P} \subseteq \mathcal{G}_A$, when $A \in \mathcal{L}(\mathcal{P})$. Now using Step 1 again, we find that $\mathcal{L}(\mathcal{P}) \subseteq \mathcal{G}_A$, and the result then follows. ■

Corollary 6.1.3 *If $\mathcal{P}$ is a π-system and $\mathcal{F}$ is a σ-field then $\sigma(\mathcal{P}) \subseteq \mathcal{F}$.*

This result can be very useful for proving that a function is measurable.

6.2 Product Measure (Δ)

Let (S_1, Σ_1, m_1) and (S_2, Σ_2, m_2) be measure spaces.

The set $S_1 \times S_2 := \{(s_1, s_2); s_1 \in S_1, s_2 \in S_2\}$ is the usual Cartesian product of sets. If $A \subseteq S_1, B \subseteq S_2$, the subset $A \times B$ of $S_1 \times S_2$ is called a *product set*.

The σ-field $\Sigma_1 \otimes \Sigma_2$ is defined to be the smallest σ-field of subsets of $S_1 \times S_2$ which contains all the product sets.

If $E \subset S_1 \times S_2$ and $x \in S_1$, then $E_x \subset S_2$ is called an *x-slice* of E where

$$E_x := \{y \in S_2; (x, y) \in E\}.$$

prop:first **Proposition 6.2.1** *If $E \in \Sigma_1 \otimes \Sigma_2$ then $E_x \in \Sigma_2$ for all $x \in S_1$.*

Corollary 6.2.2 *Let $f : S_1 \times S_2 \to \mathbb{R}$ be measurable and fix $x \in S_1$. Define $f_x : S_2 \to \mathbb{R}$ by $f_x(y) = f(x, y)$ for all $y \in S_2$. Then f_x is a measurable function.*

PROOF: For all $a \in \mathbb{R}$, we need to show that $f_x^{-1}((a, \infty)) \in \Sigma_2$. Define $E = f^{-1}((a, \infty))$. Since f is measurable, $E \in \Sigma_1 \otimes \Sigma_2$. By Proposition 6.2.1, $E_x \in \Sigma_2$. But $E_x = f_x^{-1}((a, \infty))$, and so the result follows. ■

We can similarly define y-slices of sets in $S_1 \times S_2$ and show that $f_y : S_1 \to \mathbb{R}$ is measurable, for any $y \in \mathbb{R}$, where $f_y(x) = f(x, y)$ for all $x \in S_1$.

We seek to define the product measure $m_1 \times m_2$ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$. We need to make an assumption about the measures that we are using. Let (S, Σ, m) be a measure space. We say that the measure m is σ-finite if there exists a sequence (A_n) of subsets of S with $A_n \in \Sigma$ for all $n \in \mathbb{N}$ such that $S = \bigcup_{n=1}^{\infty} A_n$ and $m(A_n) < \infty$ for all $n \in \mathbb{N}$. Clearly any finite measure (and hence all probability measures) are σ-finite. It is easy to see that Lebesgue measure on $\mathbb{R}$ is σ-finite. A measure space (S, Σ, m) is said to be σ-finite, if m is a σ-finite measure.

From now on, let (S_1, Σ_1, m_1) and (S_2, Σ_2, m_2) be σ-finite measure spaces. Let $E \in \Sigma_1 \otimes \Sigma_2$ and define $\phi_E : S_1 \to \mathbb{R}$ by

$$\phi_E(x) = m_2(E_x),$$

for all $x \in S_1$. By using the Dynkin $\pi - \lambda$ lemma, you can show that ϕ_E is measurable. Then we define product measure of E by

$$(m_1 \times m_2)(E) = \int_{S_1} \phi_E(x) dm_1(x).$$

Again using the Dynkin $\pi - \lambda$ lemma, you can show that this definition is consistent, in that

$$(m_1 \times m_2)(E) = \int_{S_2} \psi_E(y) dm_2(y),$$

where $\psi_E : S_2 \to \mathbb{R}$ is the measurable function defined by $\psi_E(y) = m_1(E_y)$ for all $y \in S_2$.

In Problem 3 you can check that if $E = A \times B$ is a product set, then

$$(m_1 \times m_2)(A \times B) = m_1(A) m_2(B).$$

It can be shown that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ and Lebesgue measure on $\mathbb{R}^2$ is precisely $\lambda \times \lambda$.

6.3 Fubini's Theorem (Δ)

We give two versions of this important result - one for general non-negative measurable functions and the other for integrable functions.

thm:fubini_1 **Theorem 6.3.1 (Fubini's Theorem 1)** *Let $f : S_1 \times S_2 \to \mathbb{R}$ be a non-negative measurable function. Then the mappings*

$$x \to \int_{S_2} f(x, y) dm_2(y),$$
$$\text{and } \; y \to \int_{S_1} f(x, y) dm_1(x),$$

are both measurable. Furthermore

$$\begin{aligned} \int_{S_1 \times S_2} f d(m_1 \times m_2) &= \int_{S_1} \left(\int_{S_2} f(x, y) dm_2(y) \right) dm_1(x) \\ &= \int_{S_2} \left(\int_{S_1} f(x, y) dm_1(x) \right) dm_2(y). \qquad (6.1) \end{aligned}$$

The proof works by first establishing the result for indicator functions, and then extending by linearity to simple functions. The next step is to take an arbitrary non-negative measurable function, approximate it by simple functions as in Theorem 2.4.1 and use the monotone convergence theorem. Note that all three integrals in (6.1) may be infinite. The next result is more useful for applications.

Theorem 6.3.2 (Fubini's Theorem 2) *Let $f : S_1 \times S_2 \to \mathbb{R}$ be an integrable function. Then the mappings*

$$x \to \int_{S_2} f(x,y)dm_2(y),$$

$$\text{and} \quad y \to \int_{S_1} f(x,y)dm_1(x),$$

are both equal (a.e.) to integrable functions. Furthermore

$$\begin{aligned}\int_{S_1\times S_2} fd(m_1 \times m_2) &= \int_{S_1}\left(\int_{S_2} f(x,y)dm_2(y)\right)dm_1(x)\\ &= \int_{S_2}\left(\int_{S_1} f(x,y)dm_1(x)\right)dm_2(y).\end{aligned} \tag{6.2}$$

The proof works by writing $f = f_+ - f_-$ and applying Theorem 6.3.1 to f_- and f_+ separately.

All of the results of this chapter extend in a straightforward way to products of finitely many measure spaces.

6.4 Exercises (Δ)

Throughout these problems (S_1, Σ_1, m_1) and (S_2, Σ_2, m_2) are measure spaces. From Problem 3 onwards they are always σ-finite.

ps:Ex

6.1 If $E, F \subset S_1 \times S_2$ and $x \in S_1$, show that

(a) $(E \cap F)_x = E_x \cap F_x$,

(b) $(E^c)_x = (E_x)^c$,

(c) $(\bigcup_{n=1}^{\infty} E_n)_x = \bigcup_{n=1}^{\infty} (E_n)_x$, where (E_n) is a sequence of subsets of $S_1 \times S_2$.

_finite_meas

6.2 If m_1 and m_2 are σ-finite measures, show that the product measure $m_1 \times m_2$ is also σ-finite.

ps:prod_meas

6.3 If $A \in \Sigma_1$ and $B \in \Sigma_2$, prove that $(m_1 \times m_2)(A \times B) = m_1(A) m_2(B)$.

:prod_meas_2

6.4 A product set $A_1 \times A_2$ is said to be *finite* if $m_i(A_i) < \infty$ for $i = 1, 2$. Show that product measure $m_1 \times m_2$ is the *unique* measure μ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ for which

$$\mu(A_1 \times A_2) = m_1(A_1) m_2(A_2),$$

for all finite product sets.

Hint: Use Dynkin's $\pi - \lambda$ lemma.

ps:prod_fgh

6.5 (a) Let $f : S_1 \to \mathbb{R}$ and $g : S_2 \to \mathbb{R}$ be measurable functions. Define $h : S_1 \times S_2 \to \mathbb{R}$ by

$$h(x, y) = f(x) g(y),$$

for all $x \in S_1, y \in S_2$. Show that h is measurable.

(b) If f and g are integrable, show that h is also integrable and that

$$\int_{S_1 \times S_2} h \, d(m_1 \times m_2) = \left(\int_{S_1} f dm_1 \right) \left(\int_{S_2} g dm_2 \right).$$

pve_sum_swap

6.6 Let $a_{ij} \geq 0$ for all $1 \leq i, j \leq \infty$. Show that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij},$$

in the sense that both double series converge to the same limit, or diverge together.

Leb_marginal

6.7 Let (S, Σ, m) be a σ-finite measure space and $f : S \to \mathbb{R}$ be a non-negative measurable function. Define $A_f = \{(x, t) \in S \times \mathbb{R}; 0 \leq t \leq f(x)\}$. Show that $A_f \in \Sigma \otimes \mathcal{B}(\mathbb{R})$ and that

$$(m \times \lambda)(A_f) = \int_S f dm.$$

ps:sinxx

6.8 Use Fubini's theorem to prove that

$$\lim_{T \to \infty} \int_0^T \frac{\sin(x)}{x} dx = \frac{\pi}{2}.$$

Hint: Write $\frac{1}{x} = \int_0^\infty e^{-xy} dy$.

fub_counterex **6.9** (a) Show that for the function $f:\mathbb{R}^2\to\mathbb{R}$, defined by $f(x,y)=\dfrac{xy}{(x^2+y^2)^2}$, the iterated integrals $\int_{-1}^1\left(\int_{-1}^1 f(x,y)dy\right)dx$ and $\int_{-1}^1\left(\int_{-1}^1 f(x,y)dx\right)dy$ exist and are equal.

(b) Show that f is not integrable over the square $-1\le x\le 1$ and $-1\le y\le 1$.

fourier_conv **6.10** ($\star$) *This problem carries on from Problems 3.18-3.22. Like those problems, it deals with properties of the Fourier transform.*

Assume that f and g are integrable functions on $\mathbb{R}$, and that g is bounded. Define the *convolution* $f*g$ of f with g by

$$(f*g)(x)=\int_{\mathbb{R}} f(x-y)g(y)dy,$$

for all $x\in\mathbb{R}$. Show that $|(f*g)(x)|<\infty$, and so $f*g$ is a well–defined function from $\mathbb{R}$ to $\mathbb{R}$. Show further that $f*g$ is both measurable and integrable, and that the Fourier transform of the convolution is the product of the Fourier transforms, i.e. that for all $y\in\mathbb{R}$,

$$\widehat{f*g}(y)=\widehat{f}(y)\widehat{g}(y).$$

Appendix A

Solutions to exercises

Chapter 1

1.1 There are $\binom{n}{r}$ subsets of size r for $0 \leq r \leq n$ and so the total number of subsets is $\sum_{r=0}^{n}\binom{n}{r} = (1+1)^2 = 2^n$. Here we used the binomial theorem $(x+y)^n = \sum_{r=0}^{n}\binom{n}{r}x^r y^{n-r}$.

1.2 To show $\Sigma_1 \cap \Sigma_2$ is a σ-field we must verify S(i) to S(iii).

S(i) Since $S \in \Sigma_1$ and $S \in \Sigma_2$, $S \in \Sigma_1 \cap \Sigma_2$.

S(ii) Suppose (A_n) is a sequence of sets in $\Sigma_1 \cap \Sigma_2$. Then $A_n \in \Sigma_1$ for all $n \in \mathbb{N}$ and so $\bigcup_{n=1}^{\infty} A_n \in \Sigma_1$. But also $A_n \in \Sigma_2$ for all $n \in \mathbb{N}$ and so $\bigcup_{n=1}^{\infty} A_n \in \Sigma_2$. Hence $\bigcup_{n=1}^{\infty} A_n \in \Sigma_1 \cap \Sigma_2$.

S(iii) If $A \in \Sigma_1 \cap \Sigma_2$, $A^c \in \Sigma_1$ and $A^c \in \Sigma_2$. Hence $A^c \in \Sigma_1 \cap \Sigma_2$.

Note that the same argument can be used to show that if $\{\Sigma_n, n \in \mathbb{N})$ are all σ-fields of subsets of S then so is $\bigcap_{n=1}^{\infty} \Sigma_n$.

$\Sigma_1 \cup \Sigma_2$ is not in general a σ-field for if $A \in \Sigma_1$ and $B \in \Sigma_2$ there is no good reason why $A \cup B \in \Sigma_1 \cup \Sigma_2$. For example let $S = \{1,2,3\}, \Sigma_1 = \{\emptyset, \{1\}, \{2,3\}, S\}, \Sigma_2 = \{\emptyset, \{2\}, \{1,3\}, S\}, A = \{1\}, B = \{2\}$. Then $A \cup B = \{1,2\}$ is neither in Σ_1 nor Σ_2.

1.3 (a) $A \cup B = [A - (A \cap B)] \cup [B - (A \cap B)] \cup (A \cap B)$ is a disjoint union, hence using finite additivity and (1.3.2)

$$m(A \cup B) = m(A - A \cap B) + m(B - A \cap B) + m(A \cap B).$$

Then

$$\begin{aligned} m(A \cup B) + m(A \cap B) &= m(A - A \cap B) + m(B - A \cap B) + 2m(A \cap B) \\ &= [m(A - A \cap B) + m(A \cap B)] \\ &+ [m(B - A \cap B) + m(A \cap B)] \\ &= m(A) + m(B), \end{aligned}$$

where we use the fact that A is the disjoint union of $A - A \cap B$ and $A \cap B$, and the analogous result for B. Note that the possibility that $m(A \cap B) = \infty$ is allowed for within this proof.

(b) $m(A \cup B) \leq m(A \cup B) + m(A \cap B) = m(A) + m(B)$ follows immediately from (a) as $m(A \cap B) \geq 0$. The general case is proved by induction. We've just established $n = 2$. Now suppose the result holds for some n. Then

$$\begin{aligned} m\left(\bigcup_{i=1}^{n+1} A_i\right) &= m\left(\bigcup_{i=1}^{n} A_i \cup A_{n+1}\right) \\ &\leq m\left(\bigcup_{i=1}^{n} A_i\right) + m(A_{n+1}) \\ &\leq \sum_{i=1}^{n} m(A_i) + m(A_{n+1}) = \sum_{i=1}^{n+1} m(A_i). \end{aligned}$$

1.4 (a) We have that $(km)(\emptyset) = km(\emptyset) = 0$ because $m(\emptyset) = 0$.

If $(A_n)_{n\in\mathbb{N}}$ is a sequence of disjoint measurable sets then

$$\sum_{n=1}^{\infty}(km)(A_n) = k\sum_{n=1}^{\infty} m(A_n) = km\left(\bigcup_{n=1}^{\infty} A_n\right) = (km)\left(\bigcup_{n=1}^{\infty} A_n\right).$$

For the second equality we use that m is σ-additive. Thus km is σ-additive.
Thus km is a measure.
If m is a finite measure, then by taking $k = \frac{1}{m(S)}$ it follows immediately that $\mathbb{P}(\cdot) = \frac{m(\cdot)}{m(S)}$ is a measure. Noting that $\mathbb{P}(S) = \frac{m(S)}{m(S)} = 1$, $\mathbb{P}$ is a probability measure.

(b) The uniform distribution m on $([a,b], \mathcal{B}([a,b]))$ is given by

$$m(A) = \frac{\lambda(A)}{b-a}$$

where λ denotes Lebesgue measure.

(c) We have that $(m+n)(\emptyset) = m(\emptyset) + n(\emptyset) = 0 + 0 = 0$.
If $(A_j)_{j\in\mathbb{N}}$ is a sequence of disjoint measurable sets then

$$\begin{aligned}
\sum_{j=1}^{\infty}(m+n)(A_j) &= \lim_{J\to\infty}\sum_{j=1}^{J} m(A_j) + n(A_j) \\
&= \lim_{J\to\infty}\sum_{j=1}^{J} m(A_j) + \sum_{j=1}^{J} n(A_j) \\
&= \sum_{j=1}^{\infty} m(A_j) + \sum_{j=1}^{\infty} n(A_j) \\
&= m\left(\bigcup_{j=1}^{\infty} A_j\right) + n\left(\bigcup_{j=1}^{\infty} A_j\right) \\
&= (m+n)\left(\bigcup_{j=1}^{\infty} A_j\right).
\end{aligned}$$

Here, the second follows because the sums are finite, and the third line follows because both series are increasing (and hence their limits both exist). The fourth line follows by σ-additivity of m and n.
Thus $m+n$ is a measure.

1.5 (a) We have $m_B(\emptyset) = m(\emptyset\cap B) = m(\emptyset) = 0$.
If $(A_n)_{n\in\mathbb{N}}$ is a sequence of disjoint measurable sets then $(A_n\cap B)_{n\in\mathbb{N}}$ are also disjoint and measurable, hence

$$\sum_{n=1}^{\infty} m_B(A_n) = \sum_{n=1}^{\infty} m(A_n\cap B) = m\left(\bigcup_{n=1}^{\infty} A_n\cap B\right) = m\left(\left(\bigcup_{n=1}^{\infty} A_n\right)\cap B\right) = m_B(\bigcup_{n=1}^{\infty} A_n).$$

Here to deduce the second equality we use the σ-additivity of m.
Thus m_B is a measure.

(b) Applying **1.4** part (a) to m_B, it is immediate that $\mathbb{P}_B$ is a probability measure.
If m itself is a probability measure, say we write $m = \mathbb{P}$, then $\mathbb{P}_B$ is the conditional distribution of $\mathbb{P}$ given that the event B occurs.

1.6 The easiest way to see that m is a measure is to first use **1.4** (a) and (c) and induction to show that if $m_1, m_2 \ldots, m_n$ are measures and $c_1, c_2, \ldots, c_n$ are non-negative numbers then $c_1m_1 + c_2m_2 + \cdots + c_nm_n$ is a measure. Now apply this with $m_j = \delta_{x_j}(1\le j\le n)$. To get a probability measure we need $\sum_{j=1}^n c_j = 1$ for then, as δ_{x_j} is a probability measure for all $1\le j\le n$, we have

$$m(S) = \sum_{j=1}^{n} c_j\delta_{x_j}(S) = \sum_{j=1}^{n} c_j = 1.$$

It's also possible to check the definition directly, but it is a little more work that way.

1.7 By definition $(a,b) \in \mathcal{B}(\mathbb{R})$. We've shown in the notes that $\{a\},\{b\} \in \mathcal{B}(\mathbb{R})$ and so by S(ii), $[a,b] = \{a\} \cup (a,b) \cup \{b\} \in \mathcal{B}(\mathbb{R})$.

1.8 (a) (i) We have that $A \cap A^c = \emptyset$ and $A \cup A^c = S$, so $m(A) + m(A^c) = m(S) = M$. Because $m(S) < \infty$ we have also that $m(A) < \infty$, hence we may subtract $m(A)$ and obtain $m(A^c) = M - m(A)$.

(ii) Let (A_n) be a decreasing sequence of sets. Then $B_n = S \setminus A_n$ defines an increasing sequence of sets, so by the first part of Theorem 1.7.1 we have $m(B_n) \to m(B)$ where $B = \cup_j B_j$.

By part (a) we have

$$m(B_n) = m(S \setminus A_n) = m(S) - m(A_n)$$
$$m(B) = m(\cup_j S \setminus A_n) = m(S \setminus \cap_j A_j) = m(S) - m(\cap_j A_j)$$

Thus $m(S) - m(A_n) \to m(S) - m(\cap_j A_j)$. Since $m(S) < \infty$ we may subtract it, and after multiplying by -1 we obtain that $m(A_n) \to m(\cap_j A_j)$.

(b) Let $S = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$ and $m = \lambda$ be Lebesgue measure on $\mathbb{R}$. Set $A_n = (-\infty, -n]$. Note that $\cap_n A_n = \emptyset$ so $\lambda(\cap_n A_n) = 0$. However, $m(A_n) = \infty$ for all n, so $m(A_n) \nrightarrow m(\cap_n A_n)$ in this case.

1.9 (a) Note that each element of Π is a subset of S. Hence Π itself is a subset of the power set $\mathcal{P}(S)$ of S. Since S is a finite set, $\mathcal{P}(S)$ is also a finite set, hence Π is also finite.

Part (b) requires you to keep a very clear head. To solve a question like this you have to explore what you have deduce from what else, with lots of thinking 'if I knew this then I would also know that' and then trying to fit a bigger picture together, connecting your start point to your desired end point. Analysis can often be like this.

(b) (i) Suppose $\Pi_i \cap \Pi_j \neq \emptyset$. Note that $\Pi_i \cap \Pi_j$ is a subset of both Π_i and Π_j.

By definition of Π, any subset of Π_i is either equal to Π_i or is equal to $\emptyset$. Since we assume that $\Pi_i \cap \Pi_j \neq \emptyset$, we therefore have $\Pi_i = \Pi_i \cap \Pi_j$. Similarly, $\Pi_j = \Pi_i \cap \Pi_j$.

Hence $\Pi_i = \Pi_j$, but this contradicts the fact that the Π_i are distinct from each other. Thus we have a contradiction and in fact we must have $\Pi_i \cap \Pi_j = \emptyset$.

(ii) By definition of Π we have $\cup_{i=1}^k \Pi_i \subseteq S$. Suppose $\cup_{i=1}^k \Pi_i \neq S$. Then $C = S \setminus \cup_{i=1}^k \Pi_i$ is a non-empty set in Σ.

Since C is disjoint from all the Π_i, we must have $C \notin \Pi$. Noting that $C \in \Sigma$, by definition of Π this implies that there is some[1] $B_1 \subset C$ such that $B_1 \neq \emptyset$.

We have that B_1 is disjoint from all the Π_i, so we must have $B_1 \notin \Pi$. Thus by the same reasoning (as we gave for C) there exists $B_2 \subset B_1$ such that $B_2 \neq \emptyset$. Iterating, we construct an infinite decreasing sequence of sets $C \supset B_1 \supset B_2 \supset B_3 \ldots$ each strictly smaller than the previous one, none of which are empty. However, this is impossible because $C \subseteq S$ is a finite set.

(iii) Let $i \in I$. So $\Pi_i \cap A \neq \emptyset$. Noting that $\Pi_i \cap A \subseteq \Pi_i$, by definition of Π we must have $\Pi_i \cap A = \Pi_i$. That is, $\Pi_i \subseteq A$. Since we have this for all $i \in I$, we have $\cup_{i\in I}\Pi_i \subseteq A$.

Now suppose that $A \setminus \cup_{i\in I}\Pi_i \neq \emptyset$. Since by (ii) we have $S = \cup_{i=1}^k \Pi_i$, and the union is disjoint by (i), this means that there is some Π_j with $j \notin I$ such that $A \cap \Pi_j \neq \emptyset$. However $A \cap \Pi_j \subseteq \Pi_j$ so by definition of Π we must have $\Pi_j \cap A = \Pi_j$. That is $\Pi_j \subseteq A$, but then we would have $j \in I$, which is a contraction.

Thus $A \setminus \cup_{i\in I}\Pi_i$ must be empty, and we conclude that $A = \cup_{i\in I}\Pi_i$.

1.10 (a) Recall that C_n is the union of 2^n disjoint closed intervals, each with length 3^{-n}, and that $C = \cap_n C_n$, with notation as in Example 1.1.1.

Suppose, for a contradiction, that $(a,b) \subseteq C$ with $a < b$. Then $(a,b) \subseteq C_n$ for all n. Choose n such that $(\frac{2}{3})^{-n} < \frac{1}{2}(b-a)$. Let us write the 2^n disjoint closed intervals making up C_n as $I_1, \ldots, I_{2^n}$. The point $c = \frac{a+b}{2}$ must fall into precisely one of these intervals, say I_j. Since I_j has length $(\frac{2}{3})^{-n}$, which is less than $\frac{1}{2}(b-a)$, we must have $I_j \subseteq (a,b)$ (draw a picture!). However, C_{n+1} does not contain all of I_j, because the middle part of I_j will be removed – so we cannot have $(a,b) \subseteq C_{n+1}$. Thus we have reached a contradiction.

(b) For a counterexample, consider a variant of the construction of the Cantor set, where instead of removing the middle thirds at stage n, we instead remove the middle $1 - e^{-1/n^2}$ (from each component of C_n). Then, by the same argument as in the proof of Lemma 1.5.5, we would have

$$\lambda(C) = \lim_n \lambda(C_n) = \lim_{n\to\infty} e^{-1}e^{-1/4}e^{-1/9}\ldots e^{-1/n^2} = \lim_{n\to\infty} \exp\left(-\sum_1^n \frac{1}{i^2}\right) = \exp\left(-\sum_1^\infty \frac{1}{n^2}\right).$$

[1] $X \subset Y$ means that $X \subseteq Y$ and $X \neq Y$ i.e. X is *strictly* smaller than the set Y

We have that $\lambda(C)$ is positive because $\sum_1^\infty \frac{1}{n^2} < \infty$.
A similar argument as in part (a) applies here, and shows that C does not contain any open intervals. The length of each interval within C_{n+1} is less than half the length of the intervals in C_n (because each interval of C_n has a middle part removed to become two intervals in C_{n+1}). Thus, by a trivial induction, each of the 2^n disjoint closed intervals in C_n has length $\leq (\frac{1}{2})^n$. You can check that we can apply the same argument as in (v), but replacing $(\frac{2}{3})^n$ with $(\frac{1}{2})^n$.

Chapter 2

2.1 (a) If $x \in A$ and $x \in B$, lhs $= 1$ and rhs$= 1 + 1 - 1 = 1$,
If $x \in A$ and $x \notin B$, lhs $= 1$ and rhs$= 1 + 0 - 0 = 1$,
If $x \notin A$ and $x \in B$, lhs $= 1$ and rhs$= 0 + 1 - 0 = 1$,
If $x \notin A$ and $x \notin B$, lhs $= 0$ and rhs$= 0 + 0 - 0 = 0$, and so we have equality of lhs and rhs in all possible cases.

(b) If $x \in A, x \notin A^c$ so lhs $= 1$ and rhs $= 1 - 0 = 1$,
if $x \notin A, x \in A^c$ so lhs $= 0$ and rhs $= 1 - 1 = 0$.

(c) Since $A = B \cup (A - B)$ and $B \cap (A - B) = \emptyset$, we can apply (a) to find that $\mathbb{1}_A = \mathbb{1}_B + \mathbb{1}_{A-B}$.

(d) The lhs and rhs are both non-zero only in the case where $x \in A$ and $x \in B$ when both lhs and rhs are 1.

For the last part, if $x \notin A$ then $x \notin A_n$ for all $n \in \mathbb{N}$ and so lhs $=$ rhs $= 0$. If $x \in A$ then $x \in A_n$ for one and only one $n \in \mathbb{N}$ and so lhs $=$ rhs $= 1$.

2.2 (a) Since for all $n \in \mathbb{N}, \sup_{k\geq n} a_k = -\inf_{k\geq n}(-a_k)$, we have

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k = \lim_{n\to\infty} \left(-\inf_{k\geq n}(-a_k)\right) = -\liminf_{n\to\infty}(-a_n).$$

(b) Since for all $n \in \mathbb{N}, \sup_{k\geq n}(a_k + b_k) \leq \sup_{k\geq n} a_k + \sup_{k\geq n} b_k$, the result is obtained similarly to (a) by taking limits on both sides.

(c) Argue as in (b) noting that the inequality is reversed for inf, or use (a) and (b) to argue that

$$\begin{aligned}\liminf_{n\to\infty}(a_n + b_n) &= -\limsup_{n\to\infty}(-a_n - b_n)\\ &\geq -\limsup_{n\to\infty}(-a_n) - \limsup_{n\to\infty}(-b_n)\\ &= \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n.\end{aligned}$$

(d) Use the fact that for all $n \in \mathbb{N}, \sup_{k\geq n}(a_k b_k) \leq \left(\sup_{k\geq n} a_k\right)\left(\sup_{k\geq n} b_k\right)$ and argue as in (b).

(e) Use the fact that for all $n \in \mathbb{N}, \inf_{k\geq n}(a_k b_k) \geq (\inf_{k\geq n} a_k)(\inf_{k\geq n} b_k)$ and argue as in (d).

(f) Since $0 \leq \liminf_{n\to\infty} |a_n| \leq \limsup_{n\to\infty} |a_n| = 0$, we must have $\liminf_{n\to\infty} |a_n| = 0$ and so $0 = \liminf_{n\to\infty} |a_n| = \limsup_{n\to\infty} |a_n|$ from which it follows that $\lim_{n\to\infty} |a_n| = 0$ and hence $\lim_{n\to\infty} a_n = 0$.

2.3 Fix $a \in \mathbb{R}$. If $a < c, f^{-1}((a,\infty)) = S \in \Sigma$ and if $a \geq c, f^{-1}((a,\infty)) = \emptyset \in \Sigma$.

2.4 If f is measurable then, writing $(a,b) = \mathbb{R} \setminus ((-\infty, a] \cup [b,\infty))$ and using the properties of pre-images,

$$f^{-1}((a,b)) = f^{-1}(\mathbb{R}) \setminus \left(f^{-1}((-\infty,a]) \cup f^{-1}([b,\infty))\right),$$

which by Theorem 2.2.1 shows that $f^{-1}((a,b)) \in \Sigma$. Note that if either a or b are infinite, the corresponding half interval above will be the empty set (which has empty pre-image).
Conversely, suppose that we have $f^{-1}((a,b)) \in \Sigma$ for all $-\infty \leq a < b \leq \infty$. Taking $b = \infty$, we have that $f^{-1}((a,\infty)) \in \Sigma$, which shows that f is measurable.

2.5 (a) For any $a \in \mathbb{R}$, we have $(f + c)^{-1}((a,\infty)) = \{x \in \mathbb{R}\,;\, f(x) + c > a\} = \{x \in \mathbb{R}\,;\, f(x) > a - c\} = f^{-1}((a - c, \infty)) \in \Sigma$ by measurability of f.

(b) First note that if $k = 0$ then $(kf)(x) = 0$ for all x, so in this case f is measurable by **2.3**.
Consider when $k > 0$. For any $a \in \mathbb{R}$, we have $(kf)^{-1}((a,\infty)) = \{x \in \mathbb{R}\,;\, kf(x) > a\} = \{x \in \mathbb{R}\,;\, f(x) > a/k\} = f^{-1}((a/k,\infty)) \in \Sigma$ by measurability of f.
For $k < 0$, we can write $kf = -(-kf)$. The function $-kf$ is measurable by the above, because $-k > 0$. Multiplying by -1 to obtain $-(-kf)$ preserves measurability by Theorem 2.4.3 , where we use that the constant function $g \equiv -1$ is measurable.
Follow-up exercise: Prove the $k < 0$ case without using Theorem 2.4.3.

2.6 $(g \circ f)^{-1}((a,\infty)) = f^{-1}(g^{-1}(a,\infty)$. Now g is Borel measurable and so $g^{-1}((a,\infty)) = A \in \mathcal{B}(\mathbb{R})$. Hence by Theorem 2.2.3, $f^{-1}(A) \in \Sigma$. So we conclude that $(g \circ f)^{-1}((a,\infty)) \in \Sigma$ and so $g \circ f$ is measurable.

If $X : \Omega \to \mathbb{R}$ is a random variable then it is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable then $g(X) = g \circ X$ is again a random variable by what we have just shown. If g is not Borel measurable then we must be wary of interpreting $g(X)$ as a random variable, unless we can directly prove that it is measurable using some other technique.

2.7 For any $a > 0$, we have

$$h^{-1}((a,\infty)) = \{x \in \mathbb{R}\,;\, f(x+y) > (a,\infty)\} = \{z - y \in \mathbb{R}\,;\, f(z) > a\} = (f^{-1}((a,\infty)))_{-y}.$$

Here we use the notation $A_y = \{a+y\,;\, a \in A\}$ from Section 1.4. Using that $A_y \in \mathcal{B}(\mathbb{R})$ whenever $A \in \mathcal{B}(\mathbb{R})$, we have that $h^{-1}((a,\infty)) \in \mathcal{B}(\mathbb{R})$, and hence h is measurable.

Alternative: Write $h = f \circ \tau_y$ where $\tau_y(x) = x + y$. The mapping τ_y is continuous and hence measurable and so h is measurable by Corollary 2.3.2.

2.8 (a) If $f(x) > 0$ then $f_+(x) = f(x)$ and $f_-(x) = 0$. If $f(x) < 0$ then $f_+(x) = 0$ and $f_-(x) = -f(x)$. If $f(x) = 0$ then $f_+(x) = f_-(x) = 0$. In all cases we have $f(x) = f_+(x) - f_-(x)$.

(b) Using the same cases as in (a), in all cases we have $|f(x)| = f_+(x) + f_-(x)$.

(c) By Theorem 2.4.1, f_+ and f_- are measurable whenever f is. By Theorem 2.4.3, the sum of measurable function is measurable, hence $|f| = f_+ + f_-$ is measurable.

2.9 If f is differentiable then it is continuous and so measurable by Corollary 2.3.2. For each $x \in \mathbb{R}, f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$. Now $x \to f(x+h)$ is measurable by Problem **2.7**, and $x \to \frac{f(x+h)-f(x)}{h}$ is measurable by Theorem 2.4.3 and Problem **2.6**(b). Finally f' is measurable by Theorem 2.4.4.

2.10 Recall that we have shown all intervals (i.e. sets of the form (a,b),$[a,b)$ and so on) are Borel sets. Another way to describe intervals is that $I \subseteq \mathbb{R}$ is an interval if, whenever $a,b \in \mathbb{R}$ and $a < c < b$ we have $c \in I$.

Suppose f is monotone increasing. Fix $c \in \mathbb{R}$ and consider $I = f^{-1}((c,\infty))$. We want to show that I is an interval. Let $a,b \in I$, so that we have $f(a), f(b) > c$, and let $a < d < b$. We have $a \le d$ so, by monotonicity of f we have $f(a) \le f(d)$. Thus $f(d) > c$, so $d \in I$. Hence I is an interval, so $I \in \mathcal{B}(\mathbb{R})$. Hence f is measurable.

2.11 (a) Let A be the set of measure zero on which f_n fails to converge to f. Then $\lim_{n\to} f_n(x) = f(x)$ for all $x \in S - A$. But then by algebra of limits $\lim_{n\to} f_n(x)^2 = f(x)^2$ for all $x \in S - A$.

(b) A be the set of measure zero on which f_n fails to converge to f and B be the set of measure zero on which g_n fails to converge to g. Now $m(A \cup B) \le m(A) + m(B) = 0$ and by algebra of limits $\lim_{n\to}(f_n(x) + g_n)(x) = f(x) + g(x)$ for all $x \in S - (A \cup B)$.

(c) This follows by writing $f_n g_n = \frac{1}{4}[(f_n + g_n)^2 - (f_n - g_n)^2]$ and using the results of (a) and (b).

2.12 (a) Let $x \in O_1 \cup O_2$. Consider if $x \in O_1$, then there is an open interval I_1 containing x. Thus I_1 is an open interval within $O_1 \cup O_2$ containing x. We can do the same for $x \in O_2$, then with $x \in I_2 \subseteq O_2$, hence $O_1 \cup O_2$ is open.

Now let $x \in O_1 \cap O_2$. Then for each $i = 1,2$ we have an open interval $I_i \subseteq O_i$ containing x. Let us write $I_1 = (a_1, b_1), I_2 = (a_2, b_2)$, and $c_1 = \max(a_1, b_1)$, $c_2 = \min(a_2, b_2)$. Then $(c_1, c_2) = I_1 \cap I_2$, and since $x \in I_1 \cap I_2$ we have $x \in (c_1, c_2)$. In particular this means $c_1 < c_2$, so $I_1 \cap I_2$ is an open interval. Also $I_1 \cap I_2 \subseteq O_1 \cap O_2$, so $O_1 \cap O_2$ is open.

(b) 1. This is true. We can use exactly the same method as in part (a): let $x \in \cup_n O_n$, and the assume $x \in O_1$ (or use O_i in place of O_1), then we have an open interval $I_1 \subseteq O_1$ containing x, then $I_1 \subseteq \cup_n O_n$, and we are done.

2. This is false. A counterexample is given by $O_n = (\frac{-1}{n}, 1 + \frac{1}{n})$, for which $\cap_n O_n = [0,1]$.

(c) Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of closed sets. Then $\mathbb{R} \setminus C_n$ is open, for each n. Using set operations we have

$$\mathbb{R} \setminus (C_1 \cup C_2) = (R \setminus C_1) \cap (\mathbb{R} \setminus C_2)$$
$$\mathbb{R} \setminus (C_1 \cap C_2) = (R \setminus C_1) \cup (\mathbb{R} \setminus C_2)$$
$$\mathbb{R} \setminus \left(\bigcup_n C_n\right) = \bigcap_n (\mathbb{R} \setminus C_n)$$
$$\mathbb{R} \setminus \left(\bigcap_n C_n\right) = \bigcup_n (\mathbb{R} \setminus C_n)$$

The first two equations combined with part (a) tell us that both the results of part (a) carry over to closed sets: both $C_1 \cap C_2$ and $C_1 \cup C_2$ are closed.

From the fourth equation, since $\mathbb{R} \setminus C_n$ is open (for all n), using (b)(i) we see that $\mathbb{R} \setminus \left(\bigcup_n C_n\right)$ is also open, hence $\bigcap_n C_n$ is closed.

However, we can't do the same for the third equation, because (b)(ii) was false. Instead, we can take complements of our counterexample in (b)(ii) to find a counterexample here, giving $C_n = \mathbb{R} \setminus (\frac{-1}{n}, 1 + \frac{1}{n}) = (-\infty, \frac{-1}{n}] \cup [1 + \frac{1}{n}, \infty)$. Then $\cup_n C_n = (-\infty, 0) \cup (1, \infty)$ which is not closed (because its complement $[0, 1]$ is not open).

2.13 (a) Its sufficient to consider the case where $x = a$. Then for any $\epsilon > 0$ and arbitrary $\delta, f(a - \delta) = 0 < f(a) + \epsilon = 1 + \epsilon$ and $f(a + \delta) = f(a) < f(a) + \epsilon = 1 + \epsilon$.

(b) Its sufficient to consider the case $x = n$ for some integer n. Again for any $\epsilon > 0$ and arbitrary $\delta, f(n - \delta) = n - 1 < f(n) + \epsilon = n + \epsilon$ and $f(n + \delta) = n < f(n) + \epsilon = n + \epsilon$.

(c) Let $U = f^{-1}((-\infty, a))$. We will show that U is open. Then it is a Borel set and f is measurable. Fix $x \in U$ and let $\epsilon = a - f(x)$. Then there exists $\delta > 0$ so that $|x - y| < \delta \Rightarrow f(y) < f(x) + \epsilon = a$ and so $y \in U$. We have shown that for each $x \in U$ there exists an open interval (of radius δ) so that if y is in this interval then $y \in U$. Hence U is open.

Chapter 3

3.1 $f = \mathbb{1}_{[-2,-1)} + 2\mathbb{1}_{[0,1)} + \mathbb{1}_{[1,2)}$.
$\int_{\mathbb{R}} f(x)dx = 1 + 2 + 1 = 4$.

3.2 If $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$, then

$$f\mathbb{1}_A = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \mathbb{1}_A = \sum_{i=1}^n c_i \mathbb{1}_{A_i \cap A} = \sum_{i=1}^n c_i \mathbb{1}_{A_i \cap A} + 0\mathbb{1}_{S \setminus A},$$

by Problem **2.1**(d). Note that $\{A \cap A_1, \dots, A \cap A_n, S \setminus A\}$ are disjoint sets, with union S.
If $f \geq 0, c_i \geq 0 (1 \leq i \leq n)$ and so $f\mathbb{1}_A \geq 0$.
If we assume that $m(A) < \infty$, then $I_A f = \sum_{i=1}^n c_i \mathbb{1} m(A_i \cap A) < \infty$, because in this case $m(A_i \cap A) < m(A) < \infty$ for all $1 \leq i \leq n$.

3.3 We have

$$\begin{aligned} f_+ &= \mathbb{1}_{[-1,0)} + 3\mathbb{1}_{[1,2)}, \\ f_- &= \mathbb{1}_{[-2,-1)} + 2\mathbb{1}_{[0,1)}, \\ \int_{\mathbb{R}} f(x)dx &= \int_{\mathbb{R}} f_+(x)dx - \int_{\mathbb{R}} f_-(x)dx = (1 + 3) - (1 + 2) = 1. \end{aligned}$$

3.4 We have already proved part (1) of Theorem 3.3.1. It remains to prove parts (2)-(4).

(2) Let $\alpha > 0$. We have

$$\begin{aligned} \int_S \alpha f dm &= \sup\left\{\int_S s\, dm\, ;\, s \text{ is simple, } 0 \leq s \leq \alpha f\right\} \\ &= \sup\left\{\int_S s\, dm\, ;\, s \text{ is simple, } 0 \leq \tfrac{1}{\alpha} s \leq f\right\} \\ &= \sup\left\{\int_S \alpha r\, dm\, ;\, r \text{ is simple, } 0 \leq r \leq f\right\} \\ &= \alpha \sup\left\{\int_S r\, dm\, ;\, r \text{ is simple, } 0 \leq r \leq f\right\} \\ &= \alpha \int_S f dm. \end{aligned}$$

Here, to deduce the third line we use that s is simple if and only if $r = \frac{1}{\alpha} s$ is simple.

(3) Since $A \subseteq B$ we have $\mathbb{1}_A \leq \mathbb{1}_B$, which means $\mathbb{1}_A f \leq \mathbb{1}_B f$. This part now follows from part (1).

(4) We have $m(A) = 0$. Suppose s is a non-negative simple function such that $0 \le s \le \mathbb{1}_A f$, and write $s = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$. Hence, for any given i, either $c_i = 0$ or we must have $A_i \subseteq A$, implying $m(A_i) = 0$. Thus $\int_S s\,dm = \sum_{i=1}^n c_i m(A_i) = 0$. Thus $\int_S f\,dm = 0$.

3.5 Imitating the proof of Lemma 3.3.2, let $A = \{x \in S; |f(x)| \ge c\}$. Then

$$\int_S f^2 dm \ge \int_A f^2 dm \ge c^2 m(A),$$

and so $m(A) \le \frac{1}{c^2} \int_S f^2 dm$.
The generalisation to $p \ge 1$ is

$$m(\{x \in S; |f(x)| \ge c\}) \le \frac{1}{c^p} \int_S |f|^p dm,$$

and it is proved similarly. Note that when p is odd, we need to replace f by $|f|$ inside the integral to ensure non-negativity.

3.6 We apply Corollary 3.3.4 to the function $g = |f|^p$, which is clearly non-negative and is measurable by Theorem 2.4.3 (compose f with the continuous function $x \mapsto |x|^p$). Thus $|f|^p = 0$ a.e. which implies that $f = 0$ a.e.

3.7 (a) We can write

$$a_n^{(N)} = \sum_{i=1}^N a_i \mathbb{1}_{\{i\}}(n).$$

Noting that $\{i\}$ is measurable and $a_i \in \mathbb{R}$, this is a simple function. By step 2 of the construction of the Lebesgue integral, noting that $\#(\{i\}) = 1$,

$$\int_{\mathbb{N}} a^{(N)}\, d\# = \sum_{i=1}^N a_i.$$

We have $a^{(N)} \to a$ pointwise as $N \to \infty$, and $0 \le a^{(N)} \le a^{(N+1)}$, so by the monotone convergence theorem

$$\int_{\mathbb{N}} a\, d\# = \lim_{N\to\infty} \int_{\mathbb{N}} a^{(N)}\, d\# = \sum_{i=1}^\infty a_i,$$

as required.

(b) We have that a is integrable if and only if $|a|$ is integrable, from part (a) occurs if and only if $\sum_n |a_n| < \infty$. That is, a is integrable if and only if it is absolutely convergent.
Lastly, writing $a = a_+ - a_-$ we have

$$\begin{aligned}
\int_{\mathbb{N}} a\, d\# &= \int_{\mathbb{N}} a_+\, d\# - \int_N a_-\, d\# \\
&= \sum_{n=1}^\infty \max(a_n, 0) - \sum_{n=1}^\infty \max(-a_n, 0) \\
&= \sum_{n=1}^\infty \max(a_n, 0) - \max(-a_n, 0) \\
&= \sum_{n=1}^\infty a_n.
\end{aligned}$$

Here, the third line follows from the second using absolute convergence, which allows us to rearrange infinite series.

3.8 (1) Using Theorem 3.3.1 (2), if $c \ge 0$,

$$\int_S cf dm = \int_S cf_+ dm - \int_S cf_- dm = c\int_S f_+ dm - c\int_S f_- dm = c\int_S f dm.$$

If $c = -1, (-f)_+ = f_-$ and $(-f)_- = f_+$ and so

$$\int_S (-f) dm = \int_S f_- dm - \int_S f_+ dm = -\left(\int_S f_+ dm - \int_S f_- dm\right) = -\int_S f dm.$$

Finally if $c < 0 (c \ne -1)$ write $c = -d$ where $d > 0$ and use the two cases we've just proved.

(3) If $f \le g$ then $g - f \ge 0$ so by Theorem 3.3.1 (1), $\int_S (g-f)dm \ge 0$. But by (1) and (2) this is equivalent to $\int_S gdm - \int_S fdm \ge 0$, i.e. $\int_S gdm \ge \int_S fdm$, as required.

3.9 (a) Noting f_+ and f_- are both non-negative, with non-negative integrals, we have

$$\left|\int_S f\,dm\right| = \left|\int_S f_+\,dm - \int_S f_-\,dm\right| \le \int_S f_+\,dm + \int_S f_-\,dm = \int_S |f|\,dm.$$

(b) By the triangle inequality we have $|f(x)+g(x)| \le |f(x)| + |g(x)|$. Thus by monotonicity and linearity (from Theorem 3.6.2) we obtain

$$\int_S |f+g|\,dm \le \int_S |f| + |g|\,dm = \int_S |f|\,dm + \int_S |g|\,dm.$$

3.10 Reflexivity is obvious as $f(x) = f(x)$ for all $x \in S$. So is symmetry, because $f(x) = g(x)$ almost everywhere if and only if $g(x) = f(x)$ almost everywhere. For transitivity, let $A = \{x \in S; f(x) \ne g(x)\}, B = \{x \in S; g(x) \ne h(x)\}$ and $C = \{x \in S; f(x) \ne h(x)\}$. Then $C \subseteq A \cup B$ and so $m(C) \le m(A) + m(B) = 0$. Thus if $f = g$ a.e. and $g = h$ a.e. we have $f = h$ a.e.

3.11 Let $x \in \mathbb{R}$ be arbitrary. Then we can find $n_0 \in \mathbb{N}$ so that $\frac{1}{n_0} < |x|$ and then for all $n \ge n_0, f_n(x) = n\mathbb{1}_{(0,1/n)}(x) = 0$. So we have proved that $\lim_{n\to} f_n(x) = 0$. But for all $n \in \mathbb{N}$

$$\int_{\mathbb{R}} |f_n(x) - 0|dx = n\int_{\mathbb{R}} \mathbb{1}_{(0,1/n)}(x)dx = n.\frac{1}{n} = 1,$$

and so we cannot find any function in the sequence that gets arbitrarily close to 0 in the $\mathcal{L}_1$ sense.

The MCT does not apply here because (f_n) is not monotone. The DCT does not apply because, if it did, then the DCT would give $\int_{\mathbb{R}} f_n \to \int_{\mathbb{R}} f = 0$ which is not true! We conclude that there is no dominating integrable function for (f_n), because all the other conditions of the DCT hold. Fatou's Lemma does apply, and would give

$$\int_{\mathbb{R}} \liminf_n f_n = \int_{\mathbb{R}} 0 \le \liminf_n \int_{\mathbb{R}} f_n = \liminf_n 1 = 1$$

which we already knew because we could calculate the integrals explicitly in this case.

3.12 First suppose that $f\mathbb{1}_A$ is integrable. Then for all $n \in \mathbb{N}, |f|\mathbb{1}_{A_n} \le |f|\mathbb{1}_A$ and so $f\mathbb{1}_{A_n}$ is integrable by monotonicity. It follows that

$$\sum_{r=1}^{n} \int_S |f|\mathbb{1}_{A_r} dm = \int_S |f|\mathbb{1}_{\bigcup_{r=1}^n A_r} dm < \infty.$$

Now $|f|\mathbb{1}_{\bigcup_{r=1}^n A_r}$ increases to $|f|\mathbb{1}_A$ as $n \to \infty$ and so by the monotone convergence theorem,

$$\sum_{r=1}^{\infty} \int_S |f|\mathbb{1}_{A_r} dm = \lim_{n\to\infty} \int_S |f|\mathbb{1}_{\bigcup_{r=1}^n A_r} dm = \int_S |f|\mathbb{1}_A dm < \infty.$$

Conversely if $f\mathbb{1}_{A_n}$ is integrable for each $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} \int_{A_n} |f|dm < \infty$, we have by Theorem 3.3.5 that

$$\begin{aligned}\int_S |f|\mathbb{1}_A dm &= \int_S |f|\mathbb{1}_{\bigcup_{n=1}^\infty A_n} dm \\ &= \sum_{n=1}^{\infty} \int_{A_n} |f|dm < \infty.\end{aligned}$$

3.13 $f - f_n \ge 0$ for all $n \in \mathbb{N}$ so by Fatou's lemma:

$$\liminf_{n\to\infty} \int_S (f - f_n)dm \ge \int_S \liminf_{n\to\infty} (f - f_n)dm.$$

$$\text{i.e. } \int_S fdm + \liminf_{n\to\infty} \int_S (-f_n)dm \ge \int_S fdm + \int_S \liminf_{n\to\infty}(-f_n)dm,$$

$$\text{and so } \liminf_{n\to\infty} -\left(\int_S f_n dm\right) \ge \int_S \liminf_{n\to\infty}(-f_n)dm.$$

Multiplying both sides by -1 reverses the inequality to yield

$$-\liminf_{n\to\infty} -\left(\int_S f_n dm\right) \le \int_S \left(-\liminf_{n\to\infty}(-f_n)\right) dm.$$

But then by definition of $\limsup_{n\to\infty}$ we have

$$\limsup_{n\to\infty}\int_S f_n dm \le \int_S \limsup_{n\to\infty} f_n dm.$$

3.14 Integrability follows easily from the facts that $|\cos(\alpha x)| \le 1$ and $|\sin(\beta x)| \le 1$ for all $x \in \mathbb{R}$. As $|\cos(x/n)f(x)| \le |f(x)|$ for all $x \in \mathbb{R}$ and f is integrable, we may use the dominated convergence theorem to deduce that

$$\lim_{n\to\infty}\int_{\mathbb{R}} \cos(x/n)f(x)dx = \int_{\mathbb{R}} \lim_{n\to\infty}\cos(x/n)f(x)dx = \int_{\mathbb{R}} f(x)dx,$$

since $\lim_{n\to\infty}\cos(x/n) = \cos(0) = 1$ for all $x \in \mathbb{R}$.

3.15 Define $f_n(x) = f(t_n, x)$ for each $n \in \mathbb{N}, x \in S$. Then $|f_n(x)| \le g(x)$ for all $x \in S$. Since g is integrable, by dominated convergence

$$\begin{aligned}\lim_{n\to\infty}\int_S f(t_n,x)dm(x) &= \int_S \lim_{n\to\infty} f_n(x)dm(x)\\ &= \int_S \lim_{n\to\infty} f(t_n,x)dm(x)\\ &= \int_S f(t,x)dm(x),\end{aligned}$$

where we used the continuity assumption (ii) in the last step.

3.16 Let (h_n) be an arbitrary sequence such that $h_n \to 0$ and define $a_{n,t}(x) = \frac{f(t_n+h,x)-f(t,x)}{h_n}$.
Since $\frac{\partial f}{\partial t}$ exists we have $a_{n,t}(x) \to \frac{\partial f}{\partial t}(x,t)$ as $n \to \infty$ for all x. By the mean value theorem there exists $\theta_n \in [0,1]$ such that $a_{n,t}(x) = \frac{\partial f}{\partial t}(t+\theta_n h, x)$, hence $|f_n(x)| \le h(x)$. Thus by dominated convergence $\int_S a_{n,t}(x)\,dm(x) \to \int_S \frac{\partial f}{\partial t}(t,x)\,dm(x)$.
By linearity of the integral we have

$$\begin{aligned}\frac{\partial}{\partial t}\int_S f(t,x)\,dm(x) &= \lim_{n\to\infty}\frac{1}{h_n}\left(\int_S f(t+h_n,x)\,dm(x) - \int_S f(t,x)\,dm(x)\right)\\ &= \lim_{n\to\infty}\int_S a_{n,t}(x)\,dm(x)\end{aligned}$$

and the result follows.

3.17 (a) For each $x \in \mathbb{R}, n \in \mathbb{N}$, the expression for $f_n(x)$ is a telescopic sum. If you begin to write it out, you see that terms cancel in pairs and you obtain

$$f_n(x) = -2xe^{-x^2} + 2(n+1)^2 xe^{-(n+1)^2x^2}.$$

Using the fact that $\lim_{N\to\infty} N^2 e^{-yN^2} = 0$, for all $y \in \mathbb{R}$ we find that

$$\lim_{n\to\infty} f_n(x) = f(x) = -2xe^{-x^2}.$$

(b) The functions f and f_n are continuous and so Riemann integrable over the closed interval $[0,a]$. We can calculate (which is left for you) that $\int_0^a f(x)dx = -2\int_0^a xe^{-x^2}dx = e^{-a^2} - 1$. But on the other hand

$$\begin{aligned}\int_0^a f_n(x)dx &= \sum_{r=1}^n \int_0^a [-2r^2xe^{-r^2x^2} + 2(r+1)^2xe^{-(r+1)^2x^2}]dx\\ &= \sum_{r=1}^n (e^{-r^2a} - e^{-(r+1)^2a})\\ &= e^{-a^2} - e^{-(n+1)^2a} \to e^{-a^2} \text{ as } n \to \infty.\end{aligned}$$

So we conclude that $\int_0^a f(x)dx \ne \lim_{n\to\infty}\int_0^a f_n(x)dx$.

3.18 Using the fact that $|e^{-ixy}| \leq 1$, we get by Theorem 3.5.1,

$$|\widehat{f}(y)| \leq \int_{\mathbb{R}} |e^{-ixy}|.|f(x)|dx \leq \int_{\mathbb{R}} |f(x)|dx < \infty.$$

For the linearity, we have

$$\begin{aligned}\widehat{af+bg}(y) &= \int_{\mathbb{R}} e^{-ixy}(af(x)+bg(x))dx \\ &= a\int_{\mathbb{R}} e^{-ixy}f(x)dx + b\int_{\mathbb{R}} e^{-ixy}g(x)dx \\ &= a\widehat{f}(y)+b\widehat{g}(y).\end{aligned}$$

3.19 $x \to \mathbb{1}_{\mathbb{Q}}(x)\cos(nx)$ is integrable as $|\mathbb{1}_{\mathbb{Q}}(x)\cos(nx)| \leq |\cos(nx)|$ for all $x \in \mathbb{R}$ and $x \to \cos(nx)$ is integrable. Similarly $x \to \mathbb{1}_{\mathbb{Q}}(x)\sin(nx)$ is integrable. So the Fourier coefficients a_n and b_n are well-defined as Lebesgue integrals. As $|\cos(nx)| \leq 1$, we have $a_n = 0$ for all $n \in \mathbb{Z}_+$ since,

$$\begin{aligned}|a_n| &\leq \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbb{1}_{\mathbb{Q}}(x)|\cos(nx)|dx \\ &\leq \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbb{1}_{\mathbb{Q}}(x)dx = 0.\end{aligned}$$

By a similar argument, $b_n = 0$ for all $n \in \mathbb{N}$. So it is possible to associate a Fourier series to $\mathbb{1}_{\mathbb{Q}}$, but this Fourier series will converges to zero!

This illustrates that pointwise convergence is not the right tool for examining convergence of Fourier series!

3.20 First observe that f_a is measurable, by Problem **2.7** (take $y = -a$ there). For integrability, we use

$$\int_{\mathbb{R}} |f_a(x)|dx = \int_{\mathbb{R}} |f(x-a)|dx = \int_{\mathbb{R}} |f(x)|dx < \infty.$$

Then

$$\widehat{f}_a(y) = \int_{\mathbb{R}} e^{-ixy}f(x-a)dx,$$

and the result follows on making a change of variable $u = x - a$.

3.21 Let $y \in \mathbb{R}$ and (y_n) be an arbitrary sequence converging to y as $n \to \infty$. We need to show that the sequence $(f(y_n))$ converges to $f(y)$. We have

$$\begin{aligned}|\widehat{f}(y_n) - \widehat{f}(y)| &= \left|\int_{\mathbb{R}} e^{-ixy_n}f(x)dx - \int_{\mathbb{R}} e^{-ixy}f(x)dx\right| \\ &\leq \int_{\mathbb{R}} |e^{-ixy_n} - e^{-ixy}|.|f(x)|dx.\end{aligned}$$

Now $|e^{-ixy_n} - e^{-ixy}| \leq |e^{-ixy_n}| + |e^{-ixy}| = 2$ and the function $x \to 2f(x)$ is integrable. Also the mapping $y \to e^{-ixy}$ is continuous, and so $\lim_{n\to\infty} |e^{-ixy_n} - e^{-ixy}| = 0$. The result follows from these two facts, and the use of Lebesgue's dominated convergence theorem.

3.22 To prove that $y \to \widehat{f}(y)$ is differentiable, we need to show that $\lim_{h\to 0}(\widehat{f}(y+h) - \widehat{f}(y))/h$ exists for each $y \in \mathbb{R}$. We have

$$\begin{aligned}\frac{\widehat{f}(y+h) - \widehat{f}(y)}{h} &= \frac{1}{h}\int_{\mathbb{R}} (e^{-ix(y+h)} - e^{-ixy})f(x)dx \\ &= \int_{\mathbb{R}} e^{-ixy}\left(\frac{e^{-ihx}-1}{h}\right)f(x)dx.\end{aligned}$$

Since $|e^{-ixy}| \leq 1$, and using the hint with $b = hx$, we get

$$\begin{aligned}\left|\frac{\widehat{f}(y+h) - \widehat{f}(y)}{h}\right| &\leq \int_{\mathbb{R}} \left|\frac{e^{-ihx}-1}{h}\right|.|f(x)|dx \\ &\leq \int_{\mathbb{R}} |x||f(x)|dx < \infty.\end{aligned}$$

Then we can use Lebesgue's dominated convergence theorem to get

$$\lim_{h\to 0}\frac{\widehat{f}(y+h)-\widehat{f}(y)}{h}=\int_{\mathbb{R}}e^{-ixy}\lim_{h\to 0}\left(\frac{e^{-ihx}-1}{h}\right)f(x)dx$$
$$=-i\int_{\mathbb{R}}e^{-ixy}xf(x)dx=-i\widehat{g}(y),$$

and the result is proved. In the last step we used

$$\lim_{h\to 0}\frac{e^{-ihx}-1}{h}=\frac{d}{dy}e^{-ixy}\bigg|_{y=0}=-ix.$$

Chapter 4

4.1 Let (a_n) be an increasing sequence that tends to ∞Define $A_n=\{\omega\in\Omega; X(\omega)\le a_n\}$. Then (A_n) increases to Ω and by Theorem 4.1.1,

$$\lim_{x\to\infty}F(x)=\lim_{n\to\infty}\mathbb{P}(A_n)=\mathbb{P}(\Omega)=1.$$

Next let $B_n=\{\omega\in\Omega; X(\omega)\le -a_n\}$. Then (B_n) decreases to $\emptyset$ and by Theorem 4.1.1,

$$\lim_{x\to-\infty}F(x)=\lim_{n\to\infty}\mathbb{P}(B_n)=\mathbb{P}(\emptyset)=0.$$

4.2 *Monotone Convergence Theorem.* Let (X_n) be an increasing sequence of non-negative random variables which converges pointwise to a random variable X, i.e. $\lim_{n\to\infty}X_n(\omega)=X(\omega)$ for all $\omega\in\Omega$ (*). Then

$$\lim_{n\to\infty}\mathbb{E}(X_n)=\mathbb{E}(X).$$

Fatou's Lemma. Let (X_n) be a sequence of non-negative random variables, then

$$\liminf_{n\to\infty}\mathbb{E}(X_n)\ge\mathbb{E}\left(\liminf_{n\to\infty}X_n\right).$$

Dominated Convergence Theorem. Let (X_n) be a sequence of random variables which converges pointwise (*) to a random variable X. Suppose that there exists an integrable, non-negative random variable Y so that $|X_n(\omega)|\le Y(\omega)$ for all $n\in\mathbb{N}$ and all $\omega\in\Omega$. Then X is integrable and

$$\lim_{n\to\infty}\mathbb{E}(X_n)=\mathbb{E}(X).$$

(*) We can in fact replace *pointwise convergence* by *convergence almost everywhere*, i.e. $\lim_{n\to\infty}X_n(\omega)=X(\omega)$ for all $\omega\in\Omega-A$, where $A\in\mathcal{F}$ and $\mathbb{P}(A)=0$.

4.3 Using the usual notation, $a\vee b=\max\{a,b\}$,

$$\mathbb{E}(\max\{X,a\})=\int_{\mathbb{R}}(x\vee a)dp_X(x).$$

Since $x\vee a\ge x$ and $x\vee a\ge a$, by monotonicity
$\int_{\mathbb{R}}(x\vee a)dp_X(x)\ge\int_{\mathbb{R}}xdp_X(x)=\mathbb{E}(X)$ and $\int_{\mathbb{R}}(x\vee a)dp_X(x)\ge\int_{\mathbb{R}}adp_X(x)=a$, and so

$$\mathbb{E}(\max\{X,a\})\ge\max\{\mathbb{E}(X),a\}.$$

4.4 We have $XY=U^2V(1-V)=0$ because $V(1-V)=0$, hence $\mathbb{E}[XY]=0$. Note that $\mathbb{E}[U]=0$. By independence of U and V, $\mathbb{E}[X]=\mathbb{E}[U]\mathbb{E}[V]=0$ and $\mathbb{E}[Y]=\mathbb{E}[U]\mathbb{E}[1-V]=0$. Hence $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$. To see that X and Y are not independent, note that $\{X=0\}=\{V=0\}$ and $\{Y=0\}=\{V=1\}$. Thus $\mathbb{P}[X=Y=0]=0$ but that $\mathbb{P}[X=0]=\mathbb{P}[Y=0]\frac{1}{2}$, so $\mathbb{P}[X=Y=0]\ne\mathbb{P}[X=0]\mathbb{P}[Y=0]$.

4.5 (a) Define $B_n=\cap_{i=1}^n A_i$. Then (B_n) is a decreasing sequence of sets and, since $\mathbb{P}$ is a finite measure, by Theorem 4.1.1 we have $\mathbb{P}[B_n]\to\mathbb{P}[\cap_{i=1}^\infty B_i]$ as $n\to\infty$. Since $\cap_{i=1}^\infty A_i=\cap_{i=1}^\infty B_i$ we thus have $\mathbb{P}[\cap_{i=1}^\infty A_i]=\lim_{n\to\infty}\mathbb{P}[\cap_{i=1}^n A_i]$. Using independence on the right hand side, we obtain

$$\mathbb{P}[\cap_{i=1}^\infty A_i]=\lim_{n\to\infty}\mathbb{P}[A_1]\mathbb{P}[A_2]\dots\mathbb{P}[A_n]=\prod_{i=1}^\infty\mathbb{P}[A_i]$$

as required. Note that the limit on the right hand side exists because $\mathbb{P}[A_1]\mathbb{P}[A_2]\dots\mathbb{P}[A_n]$ is decreasing as n increases.

(b) If $\mathbb{P}(A_n) < 1-\kappa$ for infinitely many n, where $\kappa > 0$ does not depend on n, then $\prod_{n=1}^{\infty}\mathbb{P}(A_n) = 0$ so $\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \prod_{n=1}^{\infty}\mathbb{P}(A_n)$ would hold in, for example, the case where all the (A_n) were disjoint. Disjoints events are always *dependent* (because if one occurs then all the others do not!), so clearly this 'alternative' definition is not what want.

4.6 (a) We have $\mathbb{P}[A\cap B] = \mathbb{P}[A]\cap\mathbb{P}[B]$. Noting that $A^c\cap B^c = (A\cup B)^c$, we have

$$\begin{aligned}\mathbb{P}[A^c\cap B^c] = \mathbb{P}[(A\cup B)^c] &= 1-\mathbb{P}[A\cup B]\\ &= 1-\mathbb{P}[A]-\mathbb{P}[B]+\mathbb{P}[A\cap B]\\ &= 1-\mathbb{P}[A]-\mathbb{P}[B]-\mathbb{P}[A]\mathbb{P}[B]\\ &= (1-\mathbb{P}[A])(1-\mathbb{P}[B])\\ &= \mathbb{P}[A^c]\mathbb{P}[B^c].\end{aligned}$$

Hence A^c and B^c are independent.

(b) If $A,B\in\mathcal{B}(\mathbb{R})$ and f,g are Borel measurable, then $f^{-1}(A), g^{-1}(B)\in\mathcal{B}(\mathbb{R})$ and so

$$\begin{aligned}\mathbb{P}(f(X)\in A, g(Y)\in B) &= \mathbb{P}(X\in f^{-1}(A), Y\in g^{-1}(B))\\ &= \mathbb{P}(X\in f^{-1}(A))\mathbb{P}(Y\in g^{-1}(B))\\ &= \mathbb{P}(f(X)\in A)\mathbb{P}(g(Y)\in B).\end{aligned}$$

4.7 (a) If $X=k$ and $k\in\mathbb{N}$ then

$$\sum_{n=1}^{\infty}\mathbb{1}_{\{X\geq n\}} = \sum_{n=1}^{\infty}\mathbb{1}_{\{k\geq n\}} = k = X$$

because the first k terms of the sum are 1 and the rest are 0. Since X only takes values in $\mathbb{N}$, we have

$$X = \sum_{n=1}^{\infty}\mathbb{1}_{\{X\geq n\}}.$$

By monotone convergence

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{n=1}^{\infty}\mathbb{1}_{\{X\geq n\}}\right] = \sum_{n=1}^{\infty}\mathbb{E}[\mathbb{1}_{\{X\geq n\}}] = \sum_{n=1}^{\infty}\mathbb{P}[X\geq n].$$

(b) Let $X_1 = \lfloor Y\rfloor$ and $X_2 = \lceil Y\rceil$, that is Y rounded up and down (respectively) to the nearest integer. We can apply part (a) to both X_1 and X_2, since they take values in $\mathbb{N}\cup\{0\}$.

Note that for $n\in\mathbb{N}$ we have $X_1\geq n$ if and only if $Y\geq n$. Hence

$$\sum_{n=1}^{\infty}\mathbb{P}[Y\geq n] = \sum_{n=1}^{\infty}\mathbb{P}[X_1\geq n] = \mathbb{E}[X_1]\leq\mathbb{E}[Y].$$

Here, the last line follows by monotonicity, since $X_1\leq Y$.

For X_2 we need to be slightly more careful. We have $Y\leq X_2\leq Y+1$, hence $\mathbb{P}[X_2\geq k]\leq\mathbb{P}[Y+1\geq k]$. Hence

$$\mathbb{E}[Y]\leq\mathbb{E}[X_2] = \sum_{n=1}^{\infty}\mathbb{P}[X_2\geq n]\leq\sum_{n=1}^{\infty}\mathbb{P}[Y+1\geq n] = \mathbb{1}_{\{Y\geq 0\}}+\sum_{n=1}^{\infty}\mathbb{P}[Y\geq n] = 1+\sum_{n=1}^{\infty}\mathbb{P}[Y\geq n].$$

4.8 This follows immediately from the result of Problem **3.5**, when you replace f by $X-\mu$.

4.9 (a) By linearity, the quadratic function $g(t) = \mathbb{E}(X^2)+2t\mathbb{E}(XY)+t^2\mathbb{E}(Y^2)\geq 0$ for all $t\in\mathbb{R}$. A non-negative quadratic function has at most one real root, and hence has a non-positive discriminant (i.e. $b^2-4ac\leq 0$). Hence $4\mathbb{E}(XY)^2-4\mathbb{E}(X^2)\mathbb{E}(Y^2)\leq 0$ and the result follows.

(b) Put $Y=1$ in the Cauchy-Schwarz inequality from (a) to get $\mathbb{E}(|X|)\leq\mathbb{E}(X^2)^{\frac{1}{2}}<\infty$. So X is integrable. By Problem **3.9** part (a) $|\mathbb{E}(X)|\leq\mathbb{E}(|X|)$. Combining our two inequalities gives $|\mathbb{E}(X)|^2\leq\mathbb{E}(X^2)$.

(c) If $\mathbb{E}[X^2] < \infty$ then by part (b) X is also integrable, so by linearity we have

$$\mathrm{var}(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2.$$

Hence $\mathrm{var}(X) < \infty$.

Conversely, suppose that $\mathrm{var}(X) < \infty$, and note that by assumption we also have $\mathbb{E}[X] < \infty$. We can write $X^2 = (X - \mathbb{E}[X])^2 + 2X\mathbb{E}[X] - \mathbb{E}[X]^2$ and note that all terms here are integrable by our assumptions, thus

$$\mathbb{E}[X^2] = \mathrm{var}(X) + 2\mathbb{E}[X]\mathbb{E}[X] - \mathbb{E}[X]^2 = \mathrm{var}(X) - \mathbb{E}[X]^2.$$

Hence $\mathbb{E}[X^2]$ is finite.

4.10 (a) Since $e^{-ax} \le 1$ for all $x \ge 0$ we have

$$\mathbb{E}(e^{-aX}) = \int_0^\infty e^{-ax}dp_X(x) \le \int_0^\infty dp_X(x) = 1.$$

(b)

$$\begin{aligned}
\mathbb{E}(e^{a|X|} &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{u|y|}e^{-\frac{1}{2}y^2}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 e^{-uy}e^{-\frac{1}{2}y^2}dy + \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{uy}e^{-\frac{1}{2}y^2}dy \\
&= 2.\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{uy}e^{-\frac{1}{2}y^2}dy \\
&= 2e^{\frac{1}{2}a^2}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(y-a)^2}dy \\
&= 2e^{\frac{1}{2}a^2}\frac{1}{\sqrt{2\pi}}\int_{-a}^\infty e^{-\frac{1}{2}y^2}dy \\
&= 2e^{\frac{1}{2}a^2}\mathbb{P}(X > -a) < \infty.
\end{aligned}$$

[Note that the same argument can be used to establish that X has the moment generating function $E(e^{aX}) = e^{\frac{1}{2}a^2}$.

(c) Using the fact that $e^{a|x|} = \sum_{n=0}^\infty \frac{a^n|x|^n}{n!}$ for all $x \in \mathbb{R}$ we see that for each $n \in \mathbb{N}, |x|^n \le \frac{n!}{a^n}e^{a|x|}$ and so by monotonicity:

$$\mathbb{E}(|X|^n) \le \frac{n!}{a^n}\mathbb{E}(e^{a|X|}) < \infty.$$

4.11 If f is an indicator function: $f = \mathbb{1}_A$ for some $A \in \mathcal{B}(\mathbb{R})$:

$$\int_\Omega \mathbb{1}_A(X(\omega))d\mathbb{P}(\omega) = \mathbb{P}(X \in A) = p_X(A) = \int_{\mathbb{R}} \mathbb{1}_A(x)p_X(dx),$$

and so the result holds in this case. It extends to simple functions by linearity. If f is non-negative and bounded

$$\begin{aligned}
\int_\Omega f(X(\omega))d\mathbb{P}(\omega) &= \sup\left\{\int_\Omega g(\omega)d\mathbb{P}(\omega); g \text{ simple on } \Omega, 0 \le g \le f\circ X\right\} \\
&= \sup\left\{\int_\Omega h(X(\omega))d\mathbb{P}(\omega); h \text{ simple on } \mathbb{R}, 0 \le h\circ X \le f\circ X\right\} \\
&= \sup\left\{\int_{\mathbb{R}} h(x)p_X(dx); h \text{ simple},\ 0 \le h \le f\right\} \\
&= \int_{\mathbb{R}} f(x)dp_X(x).
\end{aligned}$$

In the general case write $f = f_+ - f_-$. If f is non-negative but not necessarily bounded, the result still holds but both integrals may be (simultaneously) infinite.

Chapter 5

5.1 The sequence is assumed independent and identically distributed, which means that $\mathbb{P}[X_n \le x] = \mathbb{P}[X_1 \le x]$ for all x, and in particular $\mathbb{P}[X_n \le x] \to \mathbb{P}[X_1 \le x]$. Thus $X_n \stackrel{d}{\to} X$ in distribution.

Let $a \in (0,1]$. Since X_n only takes the value 0 and 1, $|X_n - X_1|$ only takes the values 0 and 1. Thus $\{|X_n - X| > a\} = \{|X_n - X| = 1\} = \{X_n = 1, X_1 = 0\} \cup \{X_n = 0, X_1 = 1\}$. For $n > 1$, since X_n and X_1 are independent we thus have

$$\mathbb{P}[|X_n - X| > a] = \mathbb{P}[X_n = 1, X_1 = 0] + \mathbb{P}[X_n = 0, X_1 = 1] = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$$

which does not tend to zero as $n \to \infty$. Thus X_n does not converge to X in probability.

5.2 (a) We have

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n\frac{1}{n^2} + 0\left(1 - \frac{1}{n^2}\right) = \frac{1}{n^2} \to 0$$

so $X_n \stackrel{L^1}{\to} 0$. Since $\sum \frac{1}{n^2} < \infty$, by the second Borel-Cantelli lemma we have $\mathbb{P}[X_n = n \text{ i.o.}] = 0$. Since X_n is either equal to n^2 or 0, this means that $\mathbb{P}[X_n = 0 \text{ e.v.}] = 1$. Thus $X_n \stackrel{a.s.}{\to} 0$.

(b) We have

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n\frac{1}{n} + 0\left(1 - \frac{1}{n}\right) = 1$$

which does not tend to zero, so X_n does not converge to 0 in L^1. Since $\sum \frac{1}{n} = \infty$ and the X_n are independent, by the second Borel-Cantelli lemma we have $\mathbb{P}[X_n = n \text{ i.o.}] = 1$. Thus X_n does not convergence almost surely to 0.

(c) We have

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n^2\frac{1}{n^2} + 0\left(1 - \frac{1}{n^2}\right) = 1$$

which does not tend to zero, so X_n does not converge to 0 in L^1. Since $\sum \frac{1}{n^2} < \infty$, by the second Borel-Cantelli lemma we have $\mathbb{P}[X_n = n^2 \text{ i.o.}] = 0$. Since X_n is either equal to n^2 or 0, this means that $\mathbb{P}[X_n = 0 \text{ e.v.}] = 1$. Thus $X_n \stackrel{a.s.}{\to} 0$.

(d) We have

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = \sqrt{n}\frac{1}{n} + 0\left(1 - \frac{1}{n}\right) = \frac{1}{\sqrt{n}} \to 0$$

so $X_n \stackrel{L^1}{\to} 0$. Since $\sum \frac{1}{n} = \infty$ and the X_n are independent, by the second Borel-Cantelli lemma we have $\mathbb{P}[X_n = \sqrt{n} \text{ i.o.}] = 1$. Thus X_n does not convergence almost surely to 0.

(e) In cases (a), (c) and (d) this follows from Lemma 5.2.1. For case (b), since X_n only takes the values 0 and n we have that $\{|X_n - 0| > a\} = \{X_n = n\}$ whenever $a < n$, in which case $\mathbb{P}[|X_n - 0| > a] = \mathbb{P}[X_n = n] = \frac{1}{n} \to 0$ as $n \to \infty$. Thus $X_n \stackrel{\mathbb{P}}{\to} 0$.

5.3 Let us first show that $X_n \stackrel{\mathbb{P}}{\to} 0$. Given any $\epsilon > 0$ and $c > 0$ we can find $m \in \mathbb{N}$ such that $\frac{1}{2^m c} < \epsilon$. The key point is that for $n > 2^m$ the length of the interval A_n is less than or equal to $\frac{1}{2^m}$, and since our probability measure is Lebesgue measure this gives $\mathbb{E}[1_{A_n}] \le \frac{1}{2^m}$. Hence, for all $n > 2^m$, by Markov's inequality

$$\mathbb{P}[|X_n - 0| > c] = \mathbb{P}[\mathbb{1}_{A_n} > c] \le \frac{\mathbb{E}[\mathbb{1}_{A_n}]}{c} < \frac{1}{2^m c} < \epsilon.$$

On the other hand (X_n) cannot converge to 0 almost surely since given any $n \in \mathbb{N}$, we can find $m > n$ so that A_m and A_n are disjoint, in which case for any $\omega \in \Omega$ we have $X_n(\omega) - X_m(\omega) = \mathbb{1}_{A_n}(\omega) - \mathbb{1}_{A_m}(\omega)$ is equal to either $1 - 0$ or $0 - 1$. In either case, $|X_n(\omega) - X_m(\omega)| = 1$. Since n was arbitrary and $m \ge n$, this means $X_n(\omega)$ cannot converge (to anything) as $n \to \infty$. In particular, there is no almost sure convergence to zero.

The best way to think about this question is to rewrite it in terms of probability. Lebesgue measure on $[0,1]$ is the distribution of a uniform random variable U. Then $X_n = \mathbb{1}_{A_n}$ is equal to 1 if that uniform random variable falls into A_n, and zero otherwise. Fix some sampled value for U, and then think about how the sequence X_n will behave.

5.4 Let E_m be the event that starting at the mth toss, k consecutive heads appear. Then $\mathbb{P}[E_m] = 1/2^k$. Set $A_n = E_{m+kn}$ and then the (A_n) are independent. Moreover, $\sum_{r=1}^{\infty} \mathbb{P}[A_n] = \infty$, so by the second Borel-Cantelli lemma $\mathbb{P}[A_n \text{ i.o.}] = 1$.

5.5 (a) You might reasonably think that this is obvious - if (A_n) occurs eventually then it occurs for all n after some N, and of course there are infinitely many such n so then (A_n) occurs infinitely often. Let's give a proof anyway.

Suppose $\omega \in \{A_n \text{ e.v.}\} = \bigcup_m \bigcap_{n\geq m} A_n$. Then, for at least one value of m, we have $\omega \in A_n$ for all $n \geq m$. Take any $k \in \mathbb{N}$ and pick some $n \geq \max(m,k)$. Then $\omega \in \bigcup_{i\geq k} A_i$, but this holds for all k, which implies $\omega \in \bigcap_k \bigcup_{i\geq k} A_i = \{A_i \text{ i.o.}\}$.

(b) By the laws of set algebra we have

$$\Omega \setminus \{A_n \text{ i.o.}\} = \Omega \setminus \left(\bigcap_m \bigcup_{n\geq m} A_n\right) = \bigcup_m \left(\Omega \setminus \left(\bigcup_{n\geq m} A_n\right)\right) = \bigcup_m \bigcap_{n\geq m} \Omega \setminus A_n = \{\Omega \setminus A_n \text{ e.v.}\}.$$

It follows immediately that $1 - \mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}[A_n^c \text{ e.v.}]$.

(c) Define $B_m = \cap_{n\geq m} A_n$. Note that B_m is increasing. Note that $\mathbb{P}[B_m] \leq \mathbb{P}[A_m]$ because $B_m \subseteq A_m$. Thus by Theorem 4.1.1 we have

$$\mathbb{P}[A_n \text{ e.v.}] = \mathbb{P}[\cup_m B_m] = \lim_{m\to\infty} \mathbb{P}[B_m] = \liminf_{m\to\infty} \mathbb{P}[B_m] \leq \liminf_{m\to\infty} \mathbb{P}[A_m]. \tag{A.1}$$ {eq:evbound}

In the above, we must switch from lim to lim inf before using $\mathbb{P}[B_m] \leq \mathbb{P}[A_m]$, because we cannot be sure if $\lim_n \mathbb{P}[A_n]$ exists (and in general it will not).

Using (b), we then have

$$\mathbb{P}[A_n \text{ i.o.}] = 1 - \mathbb{P}[A_n^c \text{ e.v.}] \geq 1 - \liminf_{m\to\infty} \mathbb{P}[A_m^c] = 1 - \liminf_{m\to\infty} (1 - \mathbb{P}[A_m]) = -\liminf_{m\to\infty} -\mathbb{P}[A_m] = \limsup_{m\to\infty} \mathbb{P}[A_m]. \tag{A.2}$$ {eq:iobound}

Note that $\liminf_{m\to\infty} \mathbb{P}[A_m] \leq \limsup_{m\to\infty} \mathbb{P}[A_m]$ holds automatically, from (2.1) and (2.2). Putting (A.2) and (A.2) together completes the argument.

5.6 Without loss of generality (as in the argument given for the general case) we may assume that $\mathbb{E}(X_n) = 0$ for all $n \in \mathbb{N}$. If this is not the case, we consider $X_n - \mu$ in place of X_n.

The proof proceeds in exactly the same way as when the random variables are independent, but needs the following extra calculation:

$$\begin{aligned} \text{var}(\overline{X}) &= \frac{1}{n^2}\mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_i X_j) \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}(X_i^2) \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

5.7 (a) We have both $\mathbb{P}[X_n \leq x] \to \mathbb{P}[X \leq x]$ and $\mathbb{P}[X_n \leq x] \to \mathbb{P}[Y \leq x]$, so by uniqueness of limits for real sequences, we have $\mathbb{P}[X \leq x] = \mathbb{P}[Y \leq x]$ for all $x \in \mathbb{R}$. Hence, X and Y have the same distribution (i.e. they have the same distribution functions $F_X(x) = F_Y(x)$).

(b) By definition of convergence in probability, for any $a > 0$, for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, for all $n \geq N$,

$$\mathbb{P}[|X_n - X| > a] < \epsilon \quad \text{and} \quad \mathbb{P}[|X_n - Y| > a] < \epsilon.$$

By the triangle inequality we have

$$\mathbb{P}[|X - Y| > 2a] = \mathbb{P}[|X - X_n + X_n - Y| > 2a] \leq \mathbb{P}[|X - X_n| + |X_n - Y| > 2a]. \tag{A.3}$$ {eq:ps_uniq_

If $|X - X_n| + |X_n - Y| > 2a$ then $|X - X_n| > a$ or $|X_n - Y| > a$ (or possibly both). Hence, continuing (A.3),

$$\mathbb{P}[|X - Y| > 2a] \leq \mathbb{P}[|X_n - X| > a] + \mathbb{P}[|X_n - Y| > a] \leq 2\epsilon.$$

Since this is true for any $\epsilon > 0$ and any $a > 0$, we have $\mathbb{P}[X = Y] = 1$.

5.8 Suppose $f = f_1 + if_2$ is integrable. Then both f_1 and f_2 are integrable. The integrability of $|f| = \sqrt{f_1^2 + f_2^2}$ follows immediately from the inequality $\sqrt{f_1^2 + f_2^2} \leq |f_1| + |f_2|$. For the converse use $|f_1| \leq \sqrt{f_1^2 + f_2^2}$ and $|f_2| \leq \sqrt{f_1^2 + f_2^2}$.

5.9 First suppose that we have established the case for Y, i.e. we know that $\Phi_Y(u) = e^{-\frac{1}{2}u^2}$ for all $u \in \mathbb{R}$. Then since $X = \mu + \sigma Y$, we have

$$\begin{aligned}\Phi_X(u) &= \mathbb{E}(e^{iu(\mu+\sigma Y)}) \\ &= e^{iu\mu}\mathbb{E}(e^{i(u\sigma)Y}) = e^{i\mu u - \frac{1}{2}\sigma^2 u^2},\end{aligned}$$

as was required. To establish the result for Y we write

$$\begin{aligned}\Phi_Y(u) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{iuy}e^{-\frac{1}{2}y^2}\,dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \cos(uy)e^{-\frac{1}{2}y^2}\,dy + i\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \sin(uy)e^{-\frac{1}{2}y^2}\,dy.\end{aligned}$$

As $|\cos(uy)ye^{-\frac{1}{2}u^2}| \leq |y|e^{-\frac{1}{2}y^2}$ and $|\sin(uy)ye^{-\frac{1}{2}u^2}| \leq |y|e^{-\frac{1}{2}y^2}$ and $y \to |y|e^{-\frac{1}{2}y^2}$ is integrable on $\mathbb{R}$, we may apply Problem **3.16** to deduce that $u \to \Phi_Y(u)$ is differentiable and its derivative at $u \in \mathbb{R}$ is

$$\Phi_Y'(u) = \frac{i}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{iuy}ye^{-\frac{1}{2}y^2}\,dy.$$

Now integrate by parts to find that

$$\begin{aligned}\Phi_Y'(u) &= \frac{i}{\sqrt{2\pi}}\left[-e^{iuy}e^{-\frac{1}{2}y^2}\right]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ue^{iuy}e^{-\frac{1}{2}y^2}\,dy \\ &= -u\Phi_Y(u).\end{aligned}$$

So we have the initial value problem $\dfrac{d\Phi_Y(u)}{du} = -u\Phi_Y(u)$ with initial condition, $\Phi_Y(0) = 1$ and the result follows by using the standard separation of variables technique.

5.10 First note that by Problem **4.9**, for all $1 \leq m \leq n, \mathbb{E}(|X|^m)$ is finite and so the mapping $y \to y^m$ is p_X integrable. We also have that for all $u, y \in \mathbb{R}, |i^m y^m e^{iuy}| \leq |y|^m$ Hence we can apply Problem **3.16** to differentiate up to and including n times under the integral sign to obtain

$$\frac{d^n}{du^n}\Phi_X(u) = \int_{\mathbb{R}} i^n y^n e^{iuy}\,dp_X(y).$$

Now let $u = 0$ to find that

$$\left.\frac{d^n}{du^n}\Phi_X(u)\right|_{u=0} = i^n\int_{\mathbb{R}} y^n\,dp_X(y) = i^n\mathbb{E}(X^n).$$

5.11 In this case $\mu = p$ and $\sigma = \sqrt{p(1-p)}$ and so we can write

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{1}{2}y^2}\,dy.$$

The random variable S_n is the sum of n i.i.d. Bernoulli random variables and so is binomial with mean np and variance $np(1-p)$ and so for large n it is approximately normal in the precise sense given above.

Chapter 6

6.1 (a)

$$\begin{aligned}(E\cap F)_x &= \{y \in S_2; (x,y) \in E\cap F\} \\ &= \{y \in S_2; (x,y) \in E\} \cap \{y \in S_2; (x,y) \in F\} \\ &= E_x \cap F_x.\end{aligned}$$

$$\begin{aligned}(E^c)_x &= \{y \in S_2; (x,y) \in E^c\} \\ &= \{y \in S_2; (x,y) \notin E\} \\ &= (E_x)^c.\end{aligned}$$

$$\left(\bigcup_{n=1}^{\infty} E_n\right)_x = \left\{y \in S_2; (x,y) \in \bigcup_{n=1}^{\infty} E_n\right\}$$
$$= \bigcup_{n=1}^{\infty} \{y \in S_2; (x,y) \in E_n\}$$
$$= \bigcup_{n=1}^{\infty} (E_n)_x.$$

6.2 We can write $S_1 = \bigcup_{n=1}^{\infty} A_n$ where $m_1(A_n) < \infty$ for all $n \in \mathbb{N}$ and $S_2 = \bigcup_{r=1}^{\infty} B_r$ where $m_2(B_r) < \infty$ for all $r \in \mathbb{N}$. We then have

$$S_1 \times S_2 = \bigcup_{n=1}^{\infty}\bigcup_{r=1}^{\infty} A_n \times B_r,$$

and for all $r, n \in \mathbb{N}$,

$$(m_1 \times m_2)(A_n \times B_r) = m_1(A_n)m_2(B_r) < \infty.$$

(You can, of course, write $S_1 \times S_2$ as just a single union, by using the countability of $\mathbb{N} \times \mathbb{N}$.)

6.3 Let $E = A \times B$. Then if $x \in S_1$,

$$E_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A \end{cases}$$

So $\phi_E(x) = m_2(B)\mathbb{1}_A(x)$, and hence

$$(m_1 \times m_2)(A \times B) = \int_{S_1} \phi_E(x) dm_1(x)$$
$$= m_2(B) \int_{S_1} \mathbb{1}_A(x) dm_1(x)$$
$$= m_1(A)m_2(B).$$

6.4 Suppose that μ is a measure that takes the same value as $m_1 \times m_2$ on finite product sets. Define

$$\mathcal{E} = \{E \in \Sigma_1 \otimes \Sigma_2; \mu(E) = (m_1 \times m_2)(E)\}.$$

By definition of μ, the collection $\mathcal{P}$ of all finite product sets is in $\mathcal{E}$. Since

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2),$$

it follows that $\mathcal{P}$ is a π-system. Using basic properties of measures, it is not hard to show that $\mathcal{E}$ is a λ-system (use the solution to Problem 6.1 to establish (L1)). By σ-finiteness, it follows that $\sigma(\mathcal{P}) = \Sigma_1 \otimes \Sigma_2$ and by Dynkin's $\pi - \lambda$ lemma, $\sigma(\mathcal{P}) \subseteq \mathcal{E}$. The result follows.

6.5 (a) *Method 1.* First let $f = \mathbb{1}_A$ and let $g = \mathbb{1}_B$. Since $h = \mathbb{1}_A\mathbb{1}_B = \mathbb{1}_{A\times B}$, it is clear that h is measurable in this case. Next use linearity, to extend to the case where f and g are non-negative simple functions. Next let f and g be arbitrary non-negative measurable functions. Then by Theorem 2.4.1, there is a sequence of non-negative simple functions (s_n) converging pointwise to f, and a corresponding sequence (t_m) converging pointwise to g. Taking limits as m and n go to infinity, proves that f and g are measurable in this case. Finally let f and g be arbitrary measurable functions. Write $f = f_+ - f_-$ and $g = g_+ - g_-$. Then

$$fg = (f_+g_+ + f_-g_-) - (f_-g_+ + f_+g_-),$$

is measurable as it is a sum of products of measurable functions.

Method 2. For $B \in \Sigma_2$, define $\tilde{f}_B(x,y) = f(x)\mathbb{1}_B(y)$ for all $x \in S_1, y \in S_2$. The mapping $\tilde{f} : S_1 \times S_2 \to \mathbb{R}$ is measurable since for all $a \in \mathbb{R}$, $\tilde{f}^{-1}((a,\infty)) = f^{-1}((a,\infty)) \times B \in \Sigma_1 \times \Sigma_2$. In particular, $\tilde{f}_{S_2}$ is measurable; however $\tilde{f}_{S_2}(x,y) = f(x)$ for all $x \in S_1, y \in S_2$; so $h = \tilde{f}_{S_2}\tilde{g}_{S_1}$ is the product of measurable functions, hence is measurable.

(b) Follows easily from Fubini's theorem (2).

6.6 Let m be counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Then $a(i,j) = a_{ij}$ defines a non-negative measurable function from $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where we note that $\mathcal{P}(\mathbb{N}^2) = \mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N})$. We have

$$\int_{\mathbb{N}^2} a \; d(m \times m) = \sum_{(i,j)\in\mathbb{N}^2} a_{i,j},$$

$$\int_{\mathbb{N}}\left(\int_{\mathbb{N}} a(i,j)dm(i)\right)dm(j) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{ij},$$
$$\int_{\mathbb{N}}\left(\int_{\mathbb{N}} a(i,j)dm(j)\right)dm(i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij},$$

and the result follows by Fubini's theorem 1.

6.7 (a)

$$\begin{aligned}A_f^c &= \{(x,t)\in S\times\mathbb{R}; 0\le f(x) < t\}\\ &= \bigcup_{q\in\mathbb{Q}}\{(x,t)\in S\times\mathbb{R}; 0\le f(x)<q, t\ge q\}\\ &= \bigcup_{q\in\mathbb{Q}} f^{-1}([0,q))\times[q,\infty),\end{aligned}$$

which is a countable union of measurable sets, and so is measurable. Hence $A_f = (A_f^c)^c$ is measurable.

(b) We use the definition (as a Lebesgue integral) of product measure. Fix $x\in S$. Then the x-slice $(A_f)_x$ is just the interval $[0, f(x)]$. Its Lebesgue measure is $f(x)$ and so

$$\begin{aligned}(m\times\lambda)(A_f) &= \int_S \lambda[(A_f)_x]dm(x)\\ &= \int_S f(x)dm(x).\end{aligned}$$

6.8 First fix $T>0$ and use the hint:

$$\int_0^T \frac{\sin(x)}{x}dx = \int_0^T \sin(x)\left(\int_0^\infty e^{-xy}dy\right)dx.$$

Now $f(x,y) = e^{-xy}\sin(x)$ is continuous, and so Riemann integrable (and hence Lebesgue integrable) on $[0,t]\times[0,N]$. By Fubini's theorem:

$$\begin{aligned}\int_0^T \sin(x)\left(\int_0^N e^{-xy}dy\right)dx &= \int_0^N\left(\int_0^T e^{-xy}\sin(x)dx\right)dy\\ &= -\int_0^N\left(\frac{y}{1+y^2}e^{-yT}\sin(T)\right.\\ &+ \left.\frac{1}{1+y^2}(e^{-yT}\cos(T)-1)\right)dy,\end{aligned}$$

using integration by parts.

On the other hand, $\int_0^N e^{-xy}dy = \frac{1}{x}(1-e^{-Ny})$ and so

$$\left|\int_0^N e^{-xy}dy\right| \le \frac{2}{x}.$$

Since $x\to\frac{\sin(x)}{x}$ is continuous, and hence integrable, on $[0,T]$ we can use dominated convergence to assert that

$$\begin{aligned}\int_0^T \frac{\sin(x)}{x}dx &= \int_0^T \sin(x)\left(\int_0^\infty e^{-xy}dy\right)dx\\ &= -\int_0^\infty\left(\frac{y}{1+y^2}e^{-yT}\sin(T) + \frac{1}{1+y^2}(e^{-yT}\cos(T)-1)\right)dy.\end{aligned}$$

Now use monotonicity in the first integral (since $y/1+y^2\le 1$), and dominated convergence in the second (since $|e^{-yT}\cos(T)-1|\le 2$) to deduce that

$$\lim_{T\to\infty}\int_0^T\frac{\sin(x)}{x}dx = \int_0^\infty \frac{1}{1+y^2}dy = \frac{\pi}{2}.$$

6.9 (a) Both integrals vanish by elementary calculus arguments.

(b) Let $S = \{(x,y) \in \mathbb{R}^2; -1 \leq x \leq 1, -1 \leq y \leq 1\}$ and $A = \{(x,y) \in \mathbb{R}^2; 0 \leq x \leq 1, 0 \leq y \leq 1\}$. We require $\int_S |f(x,y)|dxdy < \infty$. Note that $\int_S |f(x,y)|dxdy \geq \int_A |f(x,y)|dxdy$. Now if f were integrable over A, we could use Fubini's theorem to write it as repeated integral. But consider

$$\int_0^1 x \left(\int_0^1 \frac{y}{(x^2+y^2)^2} dy \right) dx = \frac{1}{2} \int_0^1 \left(\frac{1}{x} - \frac{x}{x^2+1} \right) dx.$$

Since $x \to \frac{1}{x}$ is not integrable over $[0,1]$, the result follows.

6.10 First observe that by Problems **2.7** and **1.5** part (a) the mapping $(x,y) \to f(x-y)g(y)$ is measurable. Let $K = \sup_{x \in \mathbb{R}} |g(x)| < \infty$, since g is bounded. Then since f is integrable

$$|(f * g)(x)| \leq \int_{\mathbb{R}} |f(x-y)|.|g(y)|dy \leq K \int_{\mathbb{R}} |f(x-y)|dy = K \int_{\mathbb{R}} |f(y)|dy < \infty.$$

We also have by Fubini's theorem

$$\begin{aligned}
\int_{\mathbb{R}} \int_{\mathbb{R}} |(f(x-y)g(y)|dydx \leq & \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x-y)|.|g(y)|dy \right) dx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x-y)|dx \right) |g(y)|dy \\
&= \int_{\mathbb{R}} |f(x)|dx \int_{\mathbb{R}} |g(y)|dy < \infty,
\end{aligned}$$

from which it follows that $f * g$ is both measurable, and integrable.

By a similar argument using Fubini's theorem, we have that

$$\begin{aligned}
\widehat{f * g}(y) &= \int_{\mathbb{R}} e^{-ixy} \int_{\mathbb{R}} f(x-z)g(z)dzdx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-iy(u+z)} f(u)du \right) g(z)dz \\
&= \int_{\mathbb{R}} e^{-iyu} f(u)du. \int_{\mathbb{R}} e^{-iyz} g(z)dz \\
&= \widehat{f}(y)\widehat{g}(y),
\end{aligned}$$

where we used that change of variable $x = u + z$.